

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model
Run on: February 25, 2003, 16:09:53 ; Search time 43 Seconds
(without alignments)
10598.073 Million cell updates/sec
Title: US-09-762-021A-1
Perfect score: 3127
Sequence: 1 ggcagagcgactgaagacca.....aaaaaaaaataangataaa 1710
Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0
Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 1816940
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries
Command line parameters:
-MODEL=frame+n2p.model -DEV=xlh
-Q=/Cyn2.1/USPTO.spool/US09762021/runat_24022003_153037_7535/app_query.fasta_1.1863
-DB=A_Geneseq_101002 -QFWT=fastan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTEXT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09762021 -CGN_1_1_41@runat_24022003_153037_7535 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : A_Geneseq_101002.*
1: /SID32/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SID32/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SID32/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SID32/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SID32/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SID32/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SID32/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SID32/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SID32/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SID32/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SID32/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID32/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID32/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID32/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID32/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID32/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID32/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID32/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID32/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID32/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID32/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID32/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID32/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Length	DB ID	Description
1	2421	77.4	593	22	AAB64372 Amino acid sequenc
2	912	29.2	179	22	AAU87627 Novel central nerv
3	912	29.2	179	22	AAU17192 Novel signal trans
4	912	29.2	179	22	AAG74084 Human colon cancer
5	912	29.2	185	22	AAU87344 Novel central nerv
6	691.5	22.1	348	22	AAM92275 Human digestive sy
7	603	19.3	121	22	AAU87343 Novel central nerv
8	602	19.3	121	22	AAU17191 Novel signal trans
9	598	19.1	134	22	AAU87626 Novel central nerv
10	598	19.1	134	22	AAU17570 Novel signal trans
11	584	18.7	535	22	AAM25595 Human protein sequ
12	579.5	18.5	723	22	AAM93529 Human polypeptide,
13	561	17.9	569	22	AAU31985 Novel human secret
14	558.5	17.9	821	14	AAR35451 Mouse eps8. Mus m
15	482	15.4	112	22	AAG75249 Human colon cancer
16	452	14.5	549	22	ABG09715 Novel human diagno
17	403	12.9	778	22	ABG60520 Drosophila melanog
18	403	12.9	778	22	ABG66160 Drosophila melanog
19	341.5	10.9	714	22	ABG71668 Drosophila melanog
20	327.5	10.5	327	22	AAB93385 Human protein sequ
21	298	9.5	764	22	ABG09702 Novel human diagno
22	271	8.7	328	22	AAG73861 Human colon cancer
23	261	8.3	75	22	AAG77134 Human colon cancer
24	256.5	8.2	280	22	AAU87645 Novel central nerv
25	256.5	8.2	280	22	ABBI0961 Human ovarian and/
26	256.5	8.2	280	22	AAU96583 Human reproductive
27	256.5	8.2	280	22	AAU17560 Novel signal trans
28	256.5	8.2	280	23	ABP41618 Human ovarian anti
29	256.5	8.2	283	22	AAU87365 Novel central nerv
30	236.5	8.2	283	22	AAU17147 Novel signal trans
31	221.5	7.2	1040	22	ABG14734 Novel human diagno
32	200.5	6.4	1078	16	AAU1704 Collagen alpha 1 (
33	200.5	6.4	1466	22	ABBS0291 Collagen type III
34	200.5	6.4	1466	23	ABBS0747 Human Tumour Endot
35	200.5	6.4	1469	22	ABG15191 Novel human diagno
36	198	6.3	1466	22	AAE02537 Porcine alpha1(iii
37	195.5	6.3	1196	13	AAU28916 Type III procollag
38	194	6.2	686	22	ABBI1397 Human secreted pro
39	193.5	6.2	1466	22	AAE02533 Bovine alpha1(iii)
40	193.5	6.2	1466	22	AAE02534 Bovine alpha1(iii)
41	191	6.1	1078	21	AAU96125 Collagen type III
42	191	6.1	1078	21	ABBS0736 Collagen type III
43	191	6.1	1078	23	ABBS09628 Amino acid sequenc
44	191	6.1	1078	23	AAE16478 Human collagen alp
45	191	6.1	1963	22	ABBS62819 Drosophila melanog

ALIGNMENTS

RESULT 1

AAB64372
ID AAB64372 standard; Protein: 593 AA.

XX AAB64372;

AC AAB64372;

DT 22-MAR-2001 (first entry)

DE Amino acid sequence of human intracellular signalling molecule INTRA4.
XX Human; intracellular signalling molecule; INTRA; immunosuppressive;
KW cytostatic; neuroprotective; nootropic; antiarteriosclerotic; cancer;
KW antiinflammatory; anti-HIV; neuroleptic; antibacterial; antifungal;
KW antiviral; antiparasitic; antihelminthic; antiparkinsonian; AIDS;
KW cell proliferative disorder; arteriosclerosis autoimmune; epilepsy;
KW inflammatory disorder; Addison's disease; gastrointestinal disorder;
KW neurological disorder; Parkinson's disease; Creutzfeldt-Jakob disease;
KW mental disorder; schizophrenia; anxiety.
XX

OS Homo sapiens.
 XX W0200077040-A2.
 XX 21-DEC-2000.
 XX 16-JUN-2000; 2000WO-US16636.
 XX 16-JUN-1999; 99US-0139566.
 PR 17-AUG-1999; 99US-0149640.
 PR 09-NOV-1999; 99US-0164417.
 XX (INCY-) INCYTE GENOMICS INC.
 XX Yue H, Tang YT, Hillman JL, Lal P, Bandman O, Baughn MR;
 PI Azimzai Y, Yang J, Reddy R, Lu DAM;
 XX WPI; 2001-025334/03.
 DR N-PSDB; AAF32641.
 XX New human intracellular signaling molecules, useful for the diagnosis,
 PT prevention and treatment of cell proliferative, autoimmune, and
 PT inflammatory, neurological, gastrointestinal, reproductive, and
 PT developmental disorders -
 XX Claim 5; Page 115-116; 192pp; English.
 PS Sequences AAF32638 - AAF32689 represent cDNA encoding human
 CC intracellular signalling molecules INTRAL - INTRAS2, represented in
 CC AAF64369 - AAF64420. Modulators of the intracellular signalling molecules
 CC of the invention exhibit immunosuppressive; cytostatic; neuroprotective;
 CC nootropic; antiarteriosclerotic; antiinflammatory; anti-HIV;
 CC antileptic; antibacterial; antifungal; antiviral; antiparasitic;
 CC antineuroleptic; and antiparkinsonian activity. INTRA polypeptides, their
 CC agonists and antagonists are useful for the treatment of a condition
 CC associated with decreased or increased expression of functional INTRA.
 CC Disorders associated with abnormal INTRA expression or activity include
 CC cell proliferative disorders e.g. arteriosclerosis and cancers;
 CC autoimmune or inflammatory disorders e.g. Addison's disease and acquired
 CC immunodeficiency syndrome (AIDS); viral, bacterial, fungal, parasitic,
 CC protozoal and helminthic infections; gastrointestinal disorders e.g.
 CC dysphagia and irritable bowel syndrome; neurological disorders e.g.
 CC epilepsy and Parkinson's disease; prion diseases e.g. Creutzfeldt-Jakob
 CC disease and mental disorders e.g. anxiety, schizophrenia and Tourette's
 CC disorder. Antibodies immuno specific for the INTRA proteins may also be
 CC useful in the diagnosis of the above disorders.
 XX Sequence 593 AA;
 SQ

Alignment Scores:
 Pred. No.: 1,74e-190 Length: 593
 Score: 2421.00 Matches: 457
 Percent Similarity: 99.78% Conservative: 0
 Best Local Similarity: 99.78% Mismatches: 1
 Query Match: 77.42% Indels: 0
 DB: 22 Gaps: 0

US-09-762-021A-1 (1-1710) x AAB64372 (1-593)

QY 2 GCAGAGGACTGAAGACCGCTCGAGAGGCTCTGGAGGAGAGCTGGAGCAAGACCT 61
 |||||
 DB 136 AlaGluArgLeuLysThrSerLeuGlnLysAlaLeuGluGluLeuGluGlnArgPro 155
 |||||
 QY 62 CGACTTGGAGGCTTCAGCCAGCCAGGACAGATGAGGGGGCTGCTATGGAAAGCCG 121
 |||||
 DB 156 ArgLeuGlyGlyLeuGlnProSerGlnAspArgTrpArgGlyProAlaMetGluArgPro 175
 |||||
 QY 122 CTCCTATGGAGGACGACGCTATCTGGAGCGGGGATCCCTCCAGAACAGCCACCCAG 181
 |||||
 DB 176 LeuProMetGluGlnAlaArgTyrLeuGluProGlyIleProGluGlnProHisGln 195
 |||||
 QY 182 AGACCCCTAGACACAGCCTCCACCATCCCAAGGCCCTGCCAGCCACACCATGGCC 241
 |||||

DB 196 ArgThrLeuGluHisSerLeuProProSerProArgProLeuProArgHisThrSerAla 215
 |||||
 QY 242 CGAGAAACCAAGTGCCTTTACTCTGCCTCTCCAAAGCGGTCTCTTCCCGGAGGACCCA 301
 |||||
 DB 216 ArgGluProSerAlaPheThrLeuProProArgSerSerSerProGluAspPro 235
 |||||
 QY 302 GAGAGGACGAGGAGTGTGAACCATGTCTAAGGGACATTGAGCTGTTCATGGGAAAG 361
 |||||
 DB 236 GluArgAspGluValLeuAsnHisValLeuArgAspIleGluLeuPheMetGlyLys 255
 |||||
 QY 362 CTGGAGAAGCCCGCAGCAAGACCAGCAGAGAAATAATTTGGGAAAAAACAAGGAC 421
 |||||
 DB 256 LeuGluLysAlaGlnAlaLysThrSerArgLysLysPheGlyLysLysAsnLysAsp 275
 |||||
 QY 422 CAGGGAGGTCTCACCAGGCAGACAGTACATTGACTGCTTCCAGAAAGATCAAGTACACTTC 481
 |||||
 DB 276 GlnGlyGlyLeuThrGlnAlaGlnIleAspCysPheGlnLysIleLysIleLysIle 295
 |||||
 QY 482 AACCTCTCTGGGAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCTCGAGCTCGTACAC 541
 |||||
 DB 296 AsnLeuLeuGlyArgLeuAlaThrTrpLeuLysGluThrSerAlaProGluLeuValHis 315
 |||||
 QY 542 ATCTCTTCAAGTCCCTGAACCTTCATCTCGGCAGGTGGCTGAGGCTGGCTAGCAGCC 601
 |||||
 DB 316 IleLeuPheLysSerLeuAsnPheIleLeuAlaArgCysProGluAlaGlyLeuAla 335
 |||||
 QY 602 CAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGCCTACAGTCTCTCTAAGC 661
 |||||
 DB 336 GlnValIleSerProLeuLeuThrProLysAlaIleAsnLeuLeuGlnSerCysLeuSer 355
 |||||
 QY 662 CCACCTGAGAGTAACCTTTGGATGGGGTTGGGCCACCTGGACCACTAGCCGGGGCGAC 721
 |||||
 DB 356 ProProGluSerAsnLeuTrpMetGlyLeuGlyProAlaTrpThrSerArgAlaAsp 375
 |||||
 QY 722 TGGACAGGCGATGAGCCCTGCCCTACCAACCCACCATTCACAGTACATGGCAACTTCCA 781
 |||||
 DB 376 TrpThrGlyAspGluProLeuProTyrGlnProThrPheSerAspAspTrpGlnLeuPro 395
 |||||
 QY 782 GAGCCCTCCAGCCCAACACCCCTTAGGATACCCAGGACCCCTGTTTCCCTCCGGGGGAAGT 841
 |||||
 DB 396 GluProSerSerGlnAlaProLeuGlyTyrGlnAspProValSerLeuArgArgGlySer 415
 |||||
 QY 842 CATAGTGTAGGAGCACCCTCACACTTTCTCAGGAGAAGACACACCAACCATGACCTCAG 901
 |||||
 DB 416 HisArgLeuGlySerThrSerHisPheProGlnGluLysThrHisAsnHisAspProGln 435
 |||||
 QY 902 CTTGGGGACCCCAACTCCAGGCCCTCCAGCCCAACCTCCAGCCAGCCCTGGAAGATG 961
 |||||
 DB 436 ProGlyAspProAsnSerArgProSerProLysProAlaGlnProAlaLeuLysMet 455
 |||||
 QY 962 CAAGTCTTGTACAGTTTGAAGCTAGGAACCCACGGGAACCTGACTGTGGTCCAGGAGAG 1021
 |||||
 DB 456 GlnValLeuTyrGluPheGluAlaArgAsnProArgGluLeuThrValValGlnGlyGlu 475
 |||||
 QY 1022 AAGCTGGAGTGTCTGGACACAGCAGCGGTGGTGGTGAAGAATGAGCGGGACGG 1081
 |||||
 DB 476 LysLeuGluValLeuAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArg 495
 |||||
 QY 1082 AGCGGTACATTCACAGCAACATCTCTGGAGCCCTTACAGCCGGGGACCCCTGGAGCCAG 1141
 |||||
 DB 496 SerGlyTyrIleProSerAsnIleLeuGluProLeuGlnProGlyThrProGlyThrGln 515
 |||||
 QY 1142 GGCAGCTCACCTCTCGGGTTCCAATGCTTCGACTTCGAGGCGCTCAAGAGGTCACA 1201
 |||||
 DB 516 GlyGlnSerProSerArgValProMetLeuArgLeuSerSerArgProGluGluValThr 535
 |||||
 QY 1202 GACTGTGCTCGAGCGAGAACCTTCTCCACTGCCACCGGTGAGGACACTTGGGTCCCTGACG 1261
 |||||
 DB 536 AspTrpLeuGlnAlaGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThr 555
 |||||
 QY 1262 GGGAGCCACTACTCTCGCATAGACCTGGGAGCTACAGATGCTATGTCCACAGGAGGCC 1321
 |||||
 DB 556 GlySerGlnLeuLeuArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAla 575
 |||||

Qy	1322	CCAGAAATCCTGTCGGGCTGGAGGTGTGCAGAGGATGCTGGGATAAGCCCT	1375
Db	576	ProArgIleLeuSerArgLeuGluaValaLargArgMetLeuGlyIleSerPro	593
RESULT 2			
AAU87627			
ID	AAU87627 standard; Protein: 179 AA.		
XX	AC	AAU87627;	
XX	DT	05-JUN-2002 (first entry)	
DE	XX	Novel central nervous system protein #537.	
XX	XX	Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;	
KW	KW	hyperproliferative disorder; neoplasm; cardiovascular disorder;	
KW	KW	cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;	
KW	KW	nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;	
KW	KW	acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;	
KW	KW	adenocarcinoma; reproductive system disorder; testicular feminisation;	
KW	KW	endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;	
KW	KW	respiratory disorder; renal disorder; kidney failure; blood disorder;	
KW	KW	myocardial infarction; wound healing; cell proliferation; skin aging;	
KW	KW	food additive; food preservative; gene therapy.	
OS	XX	Homo sapiens.	
PX	PN	WO200155318-A2.	
XX	PD	02-AUG-2001.	
PF	PF	17-JAN-2001; 2001WO-US01332.	
PR	PR	31-JAN-2000; 2000US-0179065.	
PR	PR	04-FEB-2000; 2000US-0180628.	
PR	PR	24-FEB-2000; 2000US-0184664.	
PR	PR	02-MAR-2000; 2000US-0186350.	
PR	PR	16-MAR-2000; 2000US-0189874.	
PR	PR	17-MAR-2000; 2000US-0190076.	
PR	PR	18-APR-2000; 2000US-0198123.	
PR	PR	19-MAY-2000; 2000US-0205515.	
PR	PR	07-JUN-2000; 2000US-0209467.	
PR	PR	28-JUN-2000; 2000US-0214886.	
PR	PR	30-JUN-2000; 2000US-0215135.	
PR	PR	07-JUL-2000; 2000US-0216647.	
PR	PR	07-JUL-2000; 2000US-0216880.	
PR	PR	11-JUL-2000; 2000US-0217487.	
PR	PR	11-JUL-2000; 2000US-0217496.	
PR	PR	14-JUL-2000; 2000US-0218290.	
PR	PR	26-JUL-2000; 2000US-0220963.	
PR	PR	26-JUL-2000; 2000US-0220964.	
PR	PR	14-AUG-2000; 2000US-0224518.	
PR	PR	14-AUG-2000; 2000US-0224519.	
PR	PR	14-AUG-2000; 2000US-0225213.	
PR	PR	14-AUG-2000; 2000US-0225214.	
PR	PR	14-AUG-2000; 2000US-0225266.	
PR	PR	14-AUG-2000; 2000US-0225267.	
PR	PR	14-AUG-2000; 2000US-0225268.	
PR	PR	14-AUG-2000; 2000US-0225270.	
PR	PR	14-AUG-2000; 2000US-0225447.	
PR	PR	14-AUG-2000; 2000US-0225757.	
PR	PR	14-AUG-2000; 2000US-0225758.	
PR	PR	14-AUG-2000; 2000US-0225759.	
PR	PR	18-AUG-2000; 2000US-0226279.	
PR	PR	22-AUG-2000; 2000US-0226681.	
PR	PR	22-AUG-2000; 2000US-0226868.	
PR	PR	22-AUG-2000; 2000US-0227182.	
PR	PR	23-AUG-2000; 2000US-0227009.	
PR	PR	30-AUG-2000; 2000US-0228924.	
PR	PR	01-SEP-2000; 2000US-0229287.	
PR	PR	01-SEP-2000; 2000US-0229343.	
PR	PR	01-SEP-2000; 2000US-0229344.	
PR	PR	01-SEP-2000; 2000US-0229345.	
PR	PR	05-SEP-2000; 2000US-0229509.	
PR	PR	05-SEP-2000; 2000US-0229513.	
PR	PR	06-SEP-2000; 2000US-0230437.	
PR	PR	06-SEP-2000; 2000US-0230438.	
PR	PR	08-SEP-2000; 2000US-0231242.	
PR	PR	08-SEP-2000; 2000US-0231243.	
PR	PR	08-SEP-2000; 2000US-0231244.	
PR	PR	08-SEP-2000; 2000US-0231413.	
PR	PR	08-SEP-2000; 2000US-0231414.	
PR	PR	08-SEP-2000; 2000US-0232080.	
PR	PR	08-SEP-2000; 2000US-0232081.	
PR	PR	12-SEP-2000; 2000US-0231968.	
PR	PR	14-SEP-2000; 2000US-0232397.	
PR	PR	14-SEP-2000; 2000US-0232398.	
PR	PR	14-SEP-2000; 2000US-0232399.	
PR	PR	14-SEP-2000; 2000US-0232400.	
PR	PR	14-SEP-2000; 2000US-0232401.	
PR	PR	14-SEP-2000; 2000US-0233063.	
PR	PR	14-SEP-2000; 2000US-0233064.	
PR	PR	21-SEP-2000; 2000US-0233065.	
PR	PR	21-SEP-2000; 2000US-0234223.	
PR	PR	21-SEP-2000; 2000US-0234274.	
PR	PR	25-SEP-2000; 2000US-0234997.	
PR	PR	25-SEP-2000; 2000US-0234998.	
PR	PR	26-SEP-2000; 2000US-0235484.	
PR	PR	27-SEP-2000; 2000US-0235834.	
PR	PR		

PR 17-NOV-2000; 2000US-02492115.
PR 17-NOV-2000; 2000US-02492116.
PR 17-NOV-2000; 2000US-02492117.
PR 17-NOV-2000; 2000US-02492118.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 06-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX PI Rosen CA, Barash SC, Ruben SW;
XX
XX DR WPI; 2001-581633/65.
XX N-PSDB; ABK43957.
XX
XX PT New isolated nucleic acid encoding a protein for diagnosing,
XX PT preventing, treating or ameliorating medical conditions and used as
XX PT food additives or preservatives -
XX
XX PS Claim 9; SEQ ID No 1145; 837pp; English.
XX
XX CC The invention describes an isolated nucleic acid molecule (I) encoding a
XX CC novel central nervous system protein. (I) and polypeptides (III) encoded
XX CC by (I), are used to treat a medical conditions and in diagnosis of a
XX CC pathological condition. Disorders which are diagnosed or treated include
XX CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
XX CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
XX CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
XX CC angiodysplasia, nervous system disorders e.g. Alzheimer's disease and
XX CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
XX CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
XX CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
XX CC adenocarcinomas and irritable bowel syndrome, reproductive system
XX CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
XX CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
XX CC leukaemia, disorders involving neovascularisation e.g. malignancies,
XX CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
XX CC acute kidney failure and blood related disorders e.g. myocardial
XX CC infarction. The polypeptides can also be used to aid wound healing and
XX CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
XX CC maintain organs before transplantation, for supporting cell culture of
XX CC primary tissues, to regenerate tissues and in chemotaxis. The
XX CC polypeptides can also be used as a food additive or preservative to
XX CC increase or decrease storage capabilities, fat content, lipid, protein,
XX
XX Alignment Scores:
XX Pred. No.: 1 5e-66 Length: 179
XX Score: 912.00 Matches: 179
XX Percent Similarity: 92.27% Conservative: 0
XX Best Local Similarity: 92.27% Mismatches: 0
XX Query Match: 29.17% Indels: 15
XX DB: 22 Gaps: 1
XX
XX US-09-762-021A-1 (1-1710) x AAU87627 (1-179)
XX
XX QY 796 AGCACCTTAGGATACAGGACCCCTGTTTCCTCCGGGGGAGCTCATAGTTAGGAG 855

|||||
1 SerThrLeuArgIleProGlyProCysPhePro----- 11
Db
QY 856 CACCTCACACTTTCCTCAGGAGAGACACACAAACATGACCTCAGCTGGGGACCCCAA 915
Db
12 -----Ser-GluLysThrHisAsnHisAspProGlnProGlyAspProAs 26
QY 916 CTCAGGCGCTCCAGCCCAACCTGCCAGCCAGCCCTGAAATGCAAGTCTGTAGCA 975
Db
26 nSerArgProSerProLysProAlaGlnProAlaLeuLysMetGlnValLeuTyG 46
QY 976 GTTTGAAGCTAGGAACCCACGGGAACCTGACTGTGCTCCAGGAGAGAAAGCTCGAGTTCT 1035
Db
46 uPheGluAlaArgAsnProArgGluLeuThrValValGlnGlyGluLysLeuGluVal 66
QY 1036 GGACACAGCAAGCGGTGCTGCTGCTGAAGAATGAGCGGGAGCGGCTACATTCC 1095
Db
66 uAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArgSerGlyTyIlePr 86
QY 1096 AAGCAACATCTGGAGCCCTACAGCGGGGACCCCTGGGACCCAGGCGCCAGTACCCCTC 1155
Db
86 oSerAsnIleLeuGluProLeuGlnProGlyThrProGlyThrGlnGlyGlnSerProSe 106
QY 1156 TCGGTTCCAAATGCTTCGACTTAGCTCGAGGCCCTGAAGAGGTTCACAGACTGCCTGCAGGC 1215
Db
106 rArgValProMetLeuArgLeuSerSerArgProGluGluValThrAspTrpLeuGlnAl 126
QY 1216 AGAGAACTTCTCCACTGCCAGCGGTGAGGACACTTGGGTGCCCTGACGGGAGCCAGCTACT 1275
Db
126 aGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThrGlySerGlnLeuLe 146
QY 1276 TCGATAAGACTGGGGAGCTACAGACTATGTCCACAGGAGGCGCCACCAATCTCTCTC 1335
Db
146 uArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAlaProArgIleLeuSe 166
QY 1336 CCGCTGGAGGCTGTCTAGAAGGATGCTCGGGATAAGCCCT 1375
Db
166 rArgLeuGluAlaValArgArgMetLeuGlyIleSerPro 179
RESULT 3
AAU17192
ID AAU17192 standard; Protein: 179 AA.
AC AAU17192;
XX
DT 07-NOV-2001 (first entry)
XX
DE Novel signal transduction pathway protein, Seq ID 757.
XX
KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;
KW antineoplastic; anti-HIV; antibacterial; antineoplastic; cancer;
KW immune system disorder; rheumatoid arthritis; inflammatory condition;
KW organ transplant rejection; infection; hepatitis C; blood disorder;
KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;
KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;
KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;
KW reproductive system; gastrointestinal; liver disorder; AIDS;
XX acquired immune deficiency syndrome.
XX
OS Homo sapiens.
XX
PN WO200154733-A1.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01312.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 16-MAR-2000; 2000US-0186350.
XX 02-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-01900076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.

PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 01-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465460/50.
XX N-PSDB; AAS27109.
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
XX prognosing disorders related to the proteins, including cancers, immune
XX disorders and neuronal disorders
XX Claim 1; SEQ ID No 757; 880pp; English.
XX The invention relates to novel isolated polypeptides (1), and
CC

CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.
XX

Alignment Scores:
Pred. No.: 1.5e-66 Length: 179
Score: 912.00 Matches: 179
Percent Similarity: 92.27% Conservative: 0
Best Local Similarity: 92.27% Mismatches: 0
Query Match: 29.17% Indels: 15
DB: 22 Gaps: 1

US-09-762-021A-1 (1-1710) x AAU17192 (1-179)

QY 796 AGCACCTTAGCATACAGGACCCCTTTCCCTTCGGCGGGGAAGTCATAGTTAGGGAG 855
DB 1 SerThrLeuArgIleProGlyProCysPhePro----- 11
QY 856 CACCTCACACTTTCCTCAGGAGAACACACACCAACCATGACCTCAGCCCTGGGACCCCAA 915
DB 12 -----Ser-GluLysThrHisAsnHisAspProGlnProGlyAspProAs 26
QY 916 CTCAGGCCCTCAGCCCAACCTGCCCCAGCCAGCCCTGAAATGCAAGTCTTGACGA 975
DB 26 nSerArgProSerSerProLysProAlaGlnProAlaLeuLysMetGlnValLeuTyrgl 46
QY 976 GTTTGAAGCTAGAACCCACGGAAGTACTGTGTGTCAGGAGAGAGCTGAGGTCT 1035
DB 46 upheGluAlaArgAsnProArgGluLeuThrValValGlnGlyGluLysLeuGluVal 66
QY 1036 GGACACAGCAAGCGGTGTGCTGGTGAAGATGAGCGGGACGAGCGGTACATTC 1095
DB 66 uAspHisSerLysArgTyrTrpLeuValLysAsnGluAlaGlyArgSerGlyTyrilePr 86
QY 1096 AGCACATCTCGGACCCCTCAGCCGGGGACCCCTGGGACCCAGGCGCATCACCCTC 1155
DB 86 oSerAsnIleLeuGluProGlnProGlyThrProGlyThrGlnGlyGlnSerProSe 106
QY 1156 TCGGGTTCCAAATGCTTCGACTAGCTCGAGGCTGAAGAGTCAAGAGTGCAGCTCAGGC 1215
DB 106 rArgValProMetLeuArgLeuSerSerArgProGluGluValThrAspTrpLeuGlnAl 126
QY 1216 AGAGAACTTCTCCACTGCCCGGTGAGGACACTTGGGTCCCTCAGCGGGAGCCAGCTACT 1275
DB 126 aGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThrGlySerGlnLeuLe 146
QY 1276 TGCATAGACCTGGGGACTACAGATGCTATGTCCACAGGAGGCCCCACGAATCTGTCT 1335
DB 146 uArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAlaProArgIleLeuSe 166
QY 1336 CCGGCTGGAGGTGTGCAGAGGATGCTGGGGATAGCCCT 1375
DB 166 rArgLeuGluAlaValArgMetLeuGlyIleSerPro 179

RESULT 4
AAG74084

ID AAG74084 standard; Protein; 179 AA.
XX
AC AAG74084;
XX
DT 03-SEP-2001 (first entry)
XX
DE Human colon cancer antigen protein SEQ ID NO:4848.
XX
KW Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
PA
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI; 2001-235357/24.
DR N-PSDB; AAH33515.
XX

Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
useful for preventing, diagnosing and/or treating colorectal cancers -

Claim 11; Page 6631-6632; 9803pp; English.

CC AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated P,
CC by inserting the nucleic acids into a host cell and culturing the cell,
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAH77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.

XX Sequence 179 AA:
Alignment Scores:
Pred. No.: 1.5e-66 Length: 179
Score: 912.00 Matches: 179
Percent Similarity: 92.27% Conservative: 0
Best Local Similarity: 92.27% Mismatches: 0
Query Match: 29.17% Indels: 15
DB: 22 Gaps: 1

US-09-762-021A-1 (1-1710) x AAG74084 (1-179)

QY 796 AGCACCTTAGCATACAGGACCCCTTTCCCTTCGGCGGGGAAGTCATAGTTAGGGAG 855
DB 1 SerThrLeuArgIleProGlyProCysPhePro----- 11
QY 856 CACCTCACACTTTCCTCAGGAGAACACACACCAACCATGACCTCAGCCCTGGGACCCCAA 915

```
Db 12 -----Ser-GluLysThrHisAsnHisAspProGlnProGlyAspProAs 26
Qy 916 CTCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGAATGCAAGTCTGTGACGA 975
Db 26 nSerArgProSerProLysProAlaGlnProAlaLeuLysMetGlnValLeuTyrGI 46
Qy 976 GTTTCAGCTAGGACCCACCGGAACTGACTGTGTGTCAGGAGAGAGAGTGGAGGTTCT 1035
Db 46 uPheGluAlaArgAsnProArgGluLeuThrValValGlnGlyGluLysLeuGluValLe 66
Qy 1036 GGACACAGCAAGCGGTGGTGGTGGTGAAGAATGAGCGCGGAGCGGCTACATTCC 1095
Db 66 uAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArgSerGlyTyrIlePr 86
Qy 1096 AAGCAACATCTCGAGCCCTACAGCCGGGAGCCCTGGGAGCCAGCCAGTCAACCTC 1155
Db 86 oSerAsnIleuGluProLeuGlnProGlyThrProGlyThrGlnGlyGlnSerProSe 106
Qy 1156 TCGGGTTCCAAATGCTTCGACTTAGCTGAGGCCTGAAGAGTCAAGACTGGCTGCAGGC 1215
Db 106 rArgValProMetLeuArgLeuSerSerArgProGluGluValThrAspTrpLeuGlnAl 126
Qy 1216 AGAGAACTTCTCCACTGCCAGCGTGAGACACTTGGGTCCCTGACGGGAGCCAGCTACT 1275
Db 126 aGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThrGlySerGlnLeuLe 146
Qy 1276 TCGCATAGACCTGGGAGCTACAGATGCCTATGTCACAGAGCGCCAGCCCAATCCTGTC 1335
Db 146 uArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAlaProArgIleLeuSe 166
Qy 1336 CCGGCTGGAGGCTGTCAGAGAGTCTGGGGATAAGCCCT 1375
Db 166 rArgLeuGluAlaValArgArgMetLeuGlyIleSerPro 179
RESULT 5
AAU87344
ID AAU87344 standard; Protein; 185 AA.
XX
AC AAU87344;
XX
DT 05-JUN-2002 (first entry)
XX
DE Novel central nervous system protein #254.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200155318-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01332.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216980.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226868.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229387.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
```

PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SW;
XX
XX WPI; 2001-581633/65.
DR N-PSDB; ABK43674.
DR
XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives -
XX
XX Claim 9; SEQ ID No 862; 837pp; English.
PS
XX
XX The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative

CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC anagenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Alignment Scores:

Pred. No.:	1.52e-66	Length:	185
Score:	912.00	Matches:	179
Percent Similarity:	92.27%	Conservative:	0
Best Local Similarity:	92.27%	Mismatches:	0
Query Match:	29.17%	Indels:	15
DB:	22	Gaps:	1

US-09-762-021a-1 (1-1710) x AAU87344 (1-185)

Qy	796	AGCACCCCTTAGGATACACGACCCCTGTTTCCTTCCTGGCGGGAGTCATAGGTTAGGAG	855
Db	7	SerThrLeuArgIleProGlyProCysPhePro	17
Qy	856	CACCTCACACTTTCCTCAGGAGAGACACACACCATCAGCCTCAGCCTGGGGACCCCAA	915
Db	18	-----Ser-GluLysThrHisAsnHisAspProGlnProGlyAspProAs	32
Qy	916	CTCCAGGCGCTCCAGCCCAACCTCGCCAGCCAGCCCTCGAAATGCAAGTCTTGTACGA	975
Db	32	nSerArgProSerSerProLysProAlaGlnProAlaLeuLysMetGlnValLeuTyrl	52
Qy	976	GTTTGAAGCTAGGAACCCACGGAACTGACTGTGTGCTCCAGGAGAGAGCTGGAGTTCT	1035
Db	52	uPheGluAlaArgAsnProArgGluLeuThrValValGlnGlyGluLysLeuGluVal	72
Qy	1036	GGACCACAGACAGCGGTGGTGGTGGTCAAGAAATCAGCGGGGACGGCGCTACATTCC	1095
Db	72	uAspHisSerLysArgTtrPLeuValLysAsnGluAlaGlyArgSerGlyTyrllePr	92
Qy	1096	AAGCAACATCCTGGAGCCCTACAGCGGGGACCCCTGGGACCCAGGCCAGTCAACCTC	1155
Db	92	oSerAsnIleLeuGluProLeuGlnProGlyThrProGlyThrGlnGlyGlnSerProSe	112
Qy	1156	TCGGGTTCCAATGCTTCGACTTAGCTCGAGGCCCTGAAGAGGTCACAGACTGGCTCAGGC	1215
Db	112	rArgValProMetLeuArgLeuSerSerArgProGluValThrAspTrpLeuGlnAl	132
Qy	1216	AGAGNACTTCTCCACTGCCAGCGGTGAGACACTTGGGTCCCTGAGCGGGAGCCAGCTACT	1275
Db	132	aGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThrGlySerGlnLeuLe	152
Qy	1276	TCGCATAAGACCTGGGGAGCTACAGATGCTATGTCCACAGGAGGCCCCACCAATCCTGTC	1335
Db	152	uArgIleArgProGlyGluLeuGlnMetLeuLysProGlnGluAlaProArgIleLeuSe	172
Qy	1336	CCGCTGAGGCTGTTCAGAAAGATGCTCGGGATAAGCCCT	1375
Db	172	rArgLeuGluAlaValArgMetLeuGlyIleSerPro	185
XX		RESULT 6	
XX		AAU87275	
XX		ID	AAU87275 standard; Protein: 348 AA.
XX		XX	

AC	AAM92275;		
XX		PR	14-SEP-2000; 2000US-0232400.
DT		PR	14-SEP-2000; 2000US-0232401.
XX	06-NOV-2001 (first entry)	PR	14-SEP-2000; 2000US-0233063.
XX		PR	14-SEP-2000; 2000US-0233064.
DE	Human digestive system antigen SEQ ID NO: 1624.	PR	14-SEP-2000; 2000US-0233065.
XX		PR	21-SEP-2000; 2000US-0234223.
XX	Human; digestive system antigen; gene therapy; cancer; appendicitis;	PR	21-SEP-2000; 2000US-0234274.
KW	ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;	PR	25-SEP-2000; 2000US-0234997.
KW	digestive system disorder; Meckel's diverticulum.	PR	25-SEP-2000; 2000US-0234998.
XX		PR	26-SEP-2000; 2000US-0235484.
OS	Homo sapiens.	PR	27-SEP-2000; 2000US-0235834.
XX		PR	27-SEP-2000; 2000US-0235836.
PN	WO200155314-A2.	PR	29-SEP-2000; 2000US-0236327.
XX		PR	29-SEP-2000; 2000US-0236367.
PD	02-AUG-2001.	PR	29-SEP-2000; 2000US-0236369.
XX		PR	29-SEP-2000; 2000US-0236370.
PF	17-JAN-2001; 2001WO-US01324.	PR	02-OCT-2000; 2000US-0236802.
XX		PR	02-OCT-2000; 2000US-0237037.
PR	31-JAN-2000; 2000US-0179065.	PR	02-OCT-2000; 2000US-0237038.
PR	04-FEB-2000; 2000US-0180628.	PR	02-OCT-2000; 2000US-0237039.
PR	24-FEB-2000; 2000US-0184664.	PR	02-OCT-2000; 2000US-0237039.
PR	02-MAR-2000; 2000US-0186350.	PR	02-OCT-2000; 2000US-0237040.
PR	16-MAR-2000; 2000US-0189874.	PR	13-OCT-2000; 2000US-0239935.
PR	17-MAR-2000; 2000US-0190076.	PR	13-OCT-2000; 2000US-0239937.
PR	18-APR-2000; 2000US-0198123.	PR	20-OCT-2000; 2000US-0240960.
PR	19-MAY-2000; 2000US-0205515.	PR	20-OCT-2000; 2000US-0241221.
PR	07-JUN-2000; 2000US-0209467.	PR	20-OCT-2000; 2000US-0241785.
PR	28-JUN-2000; 2000US-0214886.	PR	20-OCT-2000; 2000US-0241786.
PR	30-JUN-2000; 2000US-0215135.	PR	20-OCT-2000; 2000US-0241787.
PR	07-JUL-2000; 2000US-0216647.	PR	20-OCT-2000; 2000US-0241808.
PR	07-JUL-2000; 2000US-0216880.	PR	20-OCT-2000; 2000US-0241809.
PR	11-JUL-2000; 2000US-0217487.	PR	20-OCT-2000; 2000US-0241826.
PR	11-JUL-2000; 2000US-0217496.	PR	01-NOV-2000; 2000US-0244617.
PR	14-JUL-2000; 2000US-0218290.	PR	08-NOV-2000; 2000US-0246474.
PR	26-JUL-2000; 2000US-0220963.	PR	08-NOV-2000; 2000US-0246475.
PR	26-JUL-2000; 2000US-0220964.	PR	08-NOV-2000; 2000US-0246476.
PR	14-AUG-2000; 2000US-0224518.	PR	08-NOV-2000; 2000US-0246477.
PR	14-AUG-2000; 2000US-0224519.	PR	08-NOV-2000; 2000US-0246478.
PR	14-AUG-2000; 2000US-0225213.	PR	08-NOV-2000; 2000US-0246523.
PR	14-AUG-2000; 2000US-0225214.	PR	08-NOV-2000; 2000US-0246524.
PR	14-AUG-2000; 2000US-0225266.	PR	08-NOV-2000; 2000US-0246525.
PR	14-AUG-2000; 2000US-0225267.	PR	08-NOV-2000; 2000US-0246526.
PR	14-AUG-2000; 2000US-0225268.	PR	08-NOV-2000; 2000US-0246527.
PR	14-AUG-2000; 2000US-0225270.	PR	08-NOV-2000; 2000US-0246528.
PR	14-AUG-2000; 2000US-0225447.	PR	08-NOV-2000; 2000US-0246532.
PR	14-AUG-2000; 2000US-0225757.	PR	08-NOV-2000; 2000US-0246609.
PR	14-AUG-2000; 2000US-0225758.	PR	08-NOV-2000; 2000US-0246610.
PR	14-AUG-2000; 2000US-0225759.	PR	08-NOV-2000; 2000US-0246611.
PR	18-AUG-2000; 2000US-0226279.	PR	08-NOV-2000; 2000US-0246613.
PR	22-AUG-2000; 2000US-0226681.	PR	17-NOV-2000; 2000US-0249207.
PR	22-AUG-2000; 2000US-0226688.	PR	17-NOV-2000; 2000US-0249208.
PR	22-AUG-2000; 2000US-0227182.	PR	17-NOV-2000; 2000US-0249209.
PR	23-AUG-2000; 2000US-0227009.	PR	17-NOV-2000; 2000US-0249210.
PR	30-AUG-2000; 2000US-0228924.	PR	17-NOV-2000; 2000US-0249211.
PR	01-SEP-2000; 2000US-0229287.	PR	17-NOV-2000; 2000US-0249212.
PR	01-SEP-2000; 2000US-0229343.	PR	17-NOV-2000; 2000US-0249213.
PR	01-SEP-2000; 2000US-0229344.	PR	17-NOV-2000; 2000US-0249214.
PR	01-SEP-2000; 2000US-0229345.	PR	17-NOV-2000; 2000US-0249215.
PR	05-SEP-2000; 2000US-0229509.	PR	17-NOV-2000; 2000US-0249216.
PR	05-SEP-2000; 2000US-0229513.	PR	17-NOV-2000; 2000US-0249217.
PR	06-SEP-2000; 2000US-0230437.	PR	17-NOV-2000; 2000US-0249218.
PR	06-SEP-2000; 2000US-0230438.	PR	17-NOV-2000; 2000US-0249244.
PR	08-SEP-2000; 2000US-0231242.	PR	17-NOV-2000; 2000US-0249245.
PR	08-SEP-2000; 2000US-0231243.	PR	17-NOV-2000; 2000US-0249284.
PR	08-SEP-2000; 2000US-0231244.	PR	17-NOV-2000; 2000US-0249285.
PR	08-SEP-2000; 2000US-0231413.	PR	17-NOV-2000; 2000US-0249297.
PR	08-SEP-2000; 2000US-0231414.	PR	17-NOV-2000; 2000US-0249299.
PR	08-SEP-2000; 2000US-0232080.	PR	17-NOV-2000; 2000US-0249300.
PR	08-SEP-2000; 2000US-0232081.	PR	01-DEC-2000; 2000US-0250160.
PR	12-SEP-2000; 2000US-0231968.	PR	01-DEC-2000; 2000US-0250391.
PR	14-SEP-2000; 2000US-0232397.	PR	05-DEC-2000; 2000US-0251030.
PR	14-SEP-2000; 2000US-0232398.	PR	05-DEC-2000; 2000US-0251988.
PR	14-SEP-2000; 2000US-0232399.	PR	05-DEC-2000; 2000US-0256719.

PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.

PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-581633/65.
DR N-PSDB; ABK43673.
XX
PT New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives -
XX
PS Claim 9; SEQ ID No 861; 837pp; English.
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (II) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, endocrine disorders e.g. diabetes
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,

Alignment Scores:

Pred. No.:	3.42e-41	Length:	121
Score:	603.00	Matches:	116
Percent Similarity:	98.33%	Conservative:	2
Best Local Similarity:	96.67%	Mismatches:	2
Query Match:	19.28%	Indels:	0
DB:	22	Gaps:	0

US-09-762-021A-1 (1-1710) x AAU87343 (1-121)

Qy 1016 GGAGAGAAGCTGGAGGTTCTGGACACAGCAAGCGTGGTGGTGAAGAATGAGCG 1075
Dy 2 GlyAspSerTrpGlnValLeuAspHisSerLysArgTrpTrpValValValSAsnGluAla 21
Qy 1076 GGACGGAGCGGCTACATTCCAAACATCTCTGGAGCCCTTACAGCCGGGGACCCCTGGG 1135

Db	22	GlyArgSerGlyTyrIleProSerAsnIleLeuGluProLeuGlnProGlyThrProGly	14-AUG-2000; 2000US-0225270.
			PR 14-AUG-2000; 2000US-0225447.
QY	1136	ACCCAGGCGCAGTCACCCCTCGGGTTCCAATGCTTCGACTTAGCTCGAGCGCTGAAGAG	PR 14-AUG-2000; 2000US-0225757.
			PR 14-AUG-2000; 2000US-0225758.
Db	42	ThrGlnGlyGlnSerProSerArgValProMetLeuArgLeuSerArgProGluGlu	PR 14-AUG-2000; 2000US-0225759.
			PR 18-AUG-2000; 2000US-0226279.
QY	1196	GTCACAGACTGGCTGCAGGCAGAGAACTTCTCCACTGCCCGTGCAGGACACTGGGTCC	PR 22-AUG-2000; 2000US-0226681.
			PR 22-AUG-2000; 2000US-0226686.
Db	62	ValThrAspTrpLeuGlnAlaGluAsnPheSerThrAlaThrValArgThrLeuGlySer	PR 22-AUG-2000; 2000US-0227182.
			PR 23-AUG-2000; 2000US-0227009.
QY	1256	CTGACGGGGAGCCAGCTACTTCGCATAGACCTGGGGAGCTACAGATGCTATGTCACAG	PR 30-AUG-2000; 2000US-0228924.
			PR 01-SEP-2000; 2000US-0229287.
Db	82	LeuThrGlySerGlnLeuLeuArgIleArgProGlyGluLeuGlnMetLeuCysProGln	PR 01-SEP-2000; 2000US-0229343.
			PR 01-SEP-2000; 2000US-0229344.
QY	1316	GAGGCCCAAGAACTCTGTCGCGCTGGAGGCTGTCCAGAAAGATGCTGGGGATAAGCCCT	PR 01-SEP-2000; 2000US-0229345.
			PR 05-SEP-2000; 2000US-0229509.
Db	102	GluAlaProArgIleLeuSerArgLeuGluAlaValArgMetLeuGlyIleSerPro	PR 05-SEP-2000; 2000US-0229513.
			PR 06-SEP-2000; 2000US-0230437.
RESULT 8			PR 06-SEP-2000; 2000US-0230438.
AAU17191			PR 08-SEP-2000; 2000US-0231242.
ID AAU17191 standard; Protein; 121 AA.			PR 08-SEP-2000; 2000US-0231243.
XX AAU17191;			PR 08-SEP-2000; 2000US-0231244.
XX			PR 08-SEP-2000; 2000US-0231413.
XX			PR 08-SEP-2000; 2000US-0231414.
XX			PR 08-SEP-2000; 2000US-0232080.
XX			PR 12-SEP-2000; 2000US-0231968.
DE	07-NOV-2001 (first entry)		PR 14-SEP-2000; 2000US-0232397.
XX	Novel signal transduction pathway protein, Seq ID 756.		PR 14-SEP-2000; 2000US-0232398.
KW	Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;		PR 14-SEP-2000; 2000US-0232399.
KW	antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;		PR 14-SEP-2000; 2000US-0232400.
KW	immune system disorder; rheumatoid arthritis; inflammatory condition;		PR 14-SEP-2000; 2000US-0232401.
KW	organ transplant rejection; infection; hepatitis C; blood disorder;		PR 14-SEP-2000; 2000US-0233063.
KW	sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;		PR 14-SEP-2000; 2000US-0233064.
KW	neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;		PR 14-SEP-2000; 2000US-0233065.
KW	chromosomal abnormality; Down syndrome; ischaemia; renal disorder;		PR 21-SEP-2000; 2000US-0234223.
KW	cardiovascular; respiratory; wound healing; endocrine; Addison's disease;		PR 25-SEP-2000; 2000US-0234997.
KW	reproductive system; gastrointestinal; liver disorder; AIDS;		PR 25-SEP-2000; 2000US-0234998.
XX	acquired immune deficiency syndrome.		PR 26-SEP-2000; 2000US-0235484.
XX	Homo sapiens.		PR 27-SEP-2000; 2000US-0235834.
OS			PR 27-SEP-2000; 2000US-0235836.
PN	WO200154733-A1.		PR 29-SEP-2000; 2000US-0236327.
XX			PR 29-SEP-2000; 2000US-0236367.
XX			PR 29-SEP-2000; 2000US-0236369.
XX			PR 29-SEP-2000; 2000US-0236368.
PF	17-JAN-2001; 2001WO-US01312.		PR 29-SEP-2000; 2000US-0236370.
XX			PR 02-OCT-2000; 2000US-0236802.
PR	31-JAN-2000; 2000US-0179065.		PR 02-OCT-2000; 2000US-0237037.
PR	04-FEB-2000; 2000US-0180628.		PR 02-OCT-2000; 2000US-0237038.
PR	24-FEB-2000; 2000US-0184664.		PR 02-OCT-2000; 2000US-0237039.
PR	02-MAR-2000; 2000US-0186350.		PR 02-OCT-2000; 2000US-0237040.
PR	16-MAR-2000; 2000US-0198874.		PR 13-OCT-2000; 2000US-0239935.
PR	17-MAR-2000; 2000US-0190076.		PR 13-OCT-2000; 2000US-0239937.
PR	18-APR-2000; 2000US-0198123.		PR 20-OCT-2000; 2000US-0240960.
PR	19-MAY-2000; 2000US-0205515.		PR 20-OCT-2000; 2000US-0241221.
PR	07-JUN-2000; 2000US-0209467.		PR 20-OCT-2000; 2000US-0241785.
PR	28-JUN-2000; 2000US-0214886.		PR 20-OCT-2000; 2000US-0241786.
PR	30-JUN-2000; 2000US-0215135.		PR 20-OCT-2000; 2000US-0241787.
PR	07-JUL-2000; 2000US-0216647.		PR 20-OCT-2000; 2000US-0241808.
PR	11-JUL-2000; 2000US-0217487.		PR 20-OCT-2000; 2000US-0241809.
PR	11-JUL-2000; 2000US-0217496.		PR 20-OCT-2000; 2000US-0241826.
PR	14-JUL-2000; 2000US-0218290.		PR 01-NOV-2000; 2000US-0244617.
PR	26-JUL-2000; 2000US-0220963.		PR 08-NOV-2000; 2000US-0246474.
PR	26-JUL-2000; 2000US-0220964.		PR 08-NOV-2000; 2000US-0246475.
PR	14-AUG-2000; 2000US-0224518.		PR 08-NOV-2000; 2000US-0246476.
PR	14-AUG-2000; 2000US-0224519.		PR 08-NOV-2000; 2000US-0246477.
PR	14-AUG-2000; 2000US-0225213.		PR 08-NOV-2000; 2000US-0246478.
PR	14-AUG-2000; 2000US-0225214.		PR 08-NOV-2000; 2000US-0246523.
PR	14-AUG-2000; 2000US-0225266.		PR 08-NOV-2000; 2000US-0246524.
PR	14-AUG-2000; 2000US-0225267.		PR 08-NOV-2000; 2000US-0246525.
PR	14-AUG-2000; 2000US-0225268.		PR 08-NOV-2000; 2000US-0246526.
PR			PR 08-NOV-2000; 2000US-0246527.

PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX WPI; 2001-465460/50.
XX N-PSDB; AAS27108.
XX
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
XX prognosing disorders related to the proteins, including cancers, immune
XX disorders and neuronal disorders -
XX
XX Claim 1; SEQ ID No 756; 880pp; English.
XX
XX The invention relates to novel isolated polypeptides (I), and
XX polynucleotides (II). (I), (II) and the antibody to (I) are useful for
XX diagnosing, preventing and treating diseases including immune system
XX disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
XX disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
XX transplant rejections and graft versus host disease, infectious diseases
XX (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
XX other blood-related disorders (sickle cell anaemia), myeloproliferative
XX disorders, primary haematopoietic disorders, hyperproliferative
XX disorders (e.g. Gaucher's disease and cancer), neurodegenerative
XX disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
XX abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
XX disorders (e.g. glomerulonephritis), cardiovascular disorders
XX (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
XX wound healing, epithelial cell proliferation, endocrine disorders (e.g.
XX Addison's disease), reproductive system disorders, gastrointestinal
XX disorder (inflammatory disorders), liver disorders (cirrhosis),
XX as stimulators of B-cell responsiveness to pathogens, activators of
XX T-cells, to induce higher affinity antibodies, and as a means to induce
XX tumour proliferation in pathologies e.g. acquired immune deficiency
XX syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction

CC pathway protein, amino acid sequences of the invention.
XX
XX
XX Alignment Scores:
XX Pred. No.: 4.14e-41 Length: 121
XX Score: 602.00 Matches: 116
XX Percent Similarity: 98.33% Conservative: 2
XX Best Local Similarity: 96.67% Mismatches: 2
XX Query Match: 19.25% Indels: 0
XX DB: 22 Gaps: 0
XX
XX US-09-762-021A-1 (1-1710) x AAU17191 (1-121)
XX QY 1016 GGAGAGAAGCTGGAGGTTCTGGACACAGCAAGCGTGGTGGCTGGAAGAATGAGCG 1075
XX Db 2 GlyAsp***trpGlnValLeuAspHisSerLysArgtrpIrpLeuValLysAsnGluAla 21
XX QY 1076 GGACGAGCGGCTACATTCCAAAGCAACATCCTGGAGCCCTACAGCCGGGAGCCCTGGG 1135
XX Db 22 GlyArgSerGlyTyrlleProSerAsnIleLeuGluProLeuGlnProGlyThrProGly 41
XX QY 1136 ACCAGGGCCAGTCACCTCTCGGGTTCCAAATGCTTGCATTCGAGCTCGAGCCCTGAAGAG 1195
XX Db 42 ThrGlnGlyGlnSerProSerArgValProMetLeuArgLeuSerSerArgProGluGlu 61
XX QY 1196 GTCACAGACTGGCTGCAGGAGAGAACTTCTCCACTGCACGGTGAGGACACTTTGGGTCC 1255
XX Db 62 ValThrAspIrpLeuGlnAlaGluAsnPheserThrAlaThrValArgThrLeuGlySer 81
XX QY 1256 CTGACGGGGAGCCAGCTACTTCCGATAAGACCTGGGAGCTACAGATGCTATGTCACAG 1315
XX Db 82 LeuThrGlySerGlnLeuLeuArgIleArgProGlyGluLeuGlnMetLeuCysProGln 101
XX QY 1316 GAGCCCCCAGCAATCCTGTCGGCTGGAGGCTGTCAGAGGATGCTGGGGATGAGCCCT 1375
XX Db 102 GluAlaProArgIleLeuSerArgLeuGluAlaValArgMetLeuGlyIleSerPro 121
XX
XX RESULT 9
XX AAU87626
XX ID AAU87626 standard; Protein: 134 AA.
XX AC
XX AAU87626;
XX
XX 05-JUN-2002 (first entry)
XX
XX DE Novel central nervous system protein #536.
XX
XX Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
XX hyperproliferative disorder; neoplasm; cardiovascular disorder;
XX cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
XX nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
XX acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
XX adenocarcinoma; reproductive system disorder; testicular feminisation;
XX endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
XX respiratory disorder; renal disorder; kidney failure; blood disorder;
XX myocardial infarction; wound healing; cell proliferation; skin aging;
XX food additive; food preservative; gene therapy.
XX
XX OS Homo sapiens.
XX
XX PN WO200155318-A2.
XX
XX PD 02-AUG-2001.
XX
XX PF 17-JAN-2001; 2001WO-US01332.
XX
XX 31-JAN-2000; 2000US-0179065.
XX 04-FEB-2000; 2000US-0180628.
XX 24-FEB-2000; 2000US-0184664.
XX 02-MAR-2000; 2000US-0186350.
XX 16-MAR-2000; 2000US-0189874.
XX 17-MAR-2000; 2000US-0190076.
XX 18-APR-2000; 2000US-0198123.
XX

an cardiovascular, respiratory, wound healing; Addison's disease

KW card:

PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246509.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.

PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
XX Rosen CA, Barash SC, Ruben SM;
PI WPI; 2001-465460/50.
DR N-PSDB; AAS27487.
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders -
XX Claim 1; SEQ ID No 1135; 880pp; English.
XX The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAU17059-AAU17683 represent novel signal transduction
CC pathway protein, amino acid sequences of the invention.
XX

Alignment Scores:

Pred. No.: 9,12e-41 Length: 134
Score: 598.00 Matches: 111
Percent Similarity: 99.11% Conservative: 0
Best Local Similarity: 99.11% Mismatches: 1
Query Match: 19.12% Indels: 0
DB: 22 Gaps: 0

US-09-762-021A-1 (1-1710) x AAU17570 (1-134)

QY 827 CTTCCGCGGGGAAGTCATAGTTAGGGAGCACCTCACACTTTCCTCAGGAGAGACACAC 886
DB 14 LeuArgArgGlySerHisArgLeuGlySerThrSerHisPheProGlnGluLysThrHis 33
QY 887 AACCATGACCTCAGCCTGGGACCCCAACTCCAGGCCCTCCAGCCCAACAACTGCCAG 946
DB 34 AsnHisAspProGlnProGlyAspProAsnSerArgProSerSerProLysProAlaGln 53
QY 947 CCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCTAGGACCCACCCGGAACCTGACT 1006
DB 54 ProAlaLeuLysMetGlnValLeuTyrGluPheGluAlaArgAsnProArgGluLeuThr 73
QY 1007 GTGCTCCAGGGAGAGAAGCTGGAGGTTCTTGACACACAGCAGCGTGGTGGCTGAAG 1066
DB 74 ValValGlnGlyGluLysLeuGluValLeu***HisSerLysArgTrpTrpLeuValLys 93
QY 1067 AATAGCGGGGACGAGCGGCTACATTCACAGCAATCTCTGGAGCCCTACAGCGGGG 1126
DB 94 AsnGluAlaGlyArgSerGlyTyrIleProSerAsnIleLeuGluProLeuGlnProGly 113
QY 1127 ACCCGTGGGACCCAGGCGCAGTCACCCCTCTCGGTT 1162
DB 114 ThrProGlyThrGlnGlycInSerProSerArgVal 125

RESULT 11

AAAM25595
ID AAAM25595 standard; Protein; 535 AA.

XX AC AAAM25595;

XX DT 16-OCT-2001 (first entry)

XX DE Human protein sequence SEQ ID NO:1110.

XX KW Human; cancer; ulcer; HIV infection; human immunodeficiency virus;
XX KW antinflammatory; antirheumatic; antiarthritic; immunosuppressive;
XX KW antibacterial; endocrine; cardiac; central nervous system; virucide;
XX KW anti-HIV; fungicide; antimutagen; cardiovascular; antianemic; anaemia;
XX KW antiagregant; haemostatic; vulnery; antiulcer; osteopathic; eczema;
XX KW dermatological; antiallergic; antidiabetic; antidiabetic; cytostatic;
XX KW neuroprotective; antidepressant; antiparkinsonian; infection;
XX KW immunostimulant; gene therapy; antisense therapy; vaccine; inflammation;
XX KW antianaphylactic; rheumatoid arthritis; septic shock; pancreatitis;
XX KW cardiac dysfunction; neuropathology; cardiac anaphylaxis; autoimmunity;
XX KW genetic disease; haematopoietic disorder; platelet disorder; asthma;
XX KW thrombocytopaenia; osteoporosis; severe combined immunodeficiency;
XX KW allergic rhinitis; diabetes; multiple sclerosis; depression;
XX KW Alzheimer's disease; Parkinson's disease; neurodegenerative disorder;
XX KW neurological disorder.

XX OS Homo sapiens.

XX PN WO200153455-A2.

XX PD 26-JUL-2001.

XX PF 22-DEC-2000; 2000WO-US35017.

XX PR 23-DEC-1999; 99US-0471275.

XX PR 21-JAN-2000; 2000US-0488725.

XX PR 25-APR-2000; 2000US-0552317.

XX PA (HYSE-) HYSEQ INC.

XX PI Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-457603/49.

XX DR N-PSDB; AAH99536.

XX PT Isolated human polynucleotides encoding polypeptides, useful for the
XX PT treatment and diagnosis of e.g. cancer, ulcers and HIV infection -

XX PS Claim 20; Page 228; 1217pp; English.

XX CC AAH99166 to AAH99904 encode the human proteins given in AAAM25225 to
XX CC AAAM25963. The proteins can have activities based on the tissues and
XX CC cells they are expressed in, such as: antinflammatory; antirheumatic;
XX CC antiarthritic; immunosuppressive; antibacterial; endocrine; cardiac;
XX CC central nervous system; virucide; anti-HIV; fungicide; antimutagen;
XX CC cardiovascular; antianemic; antiagregant; haemostatic; vulnery;
XX CC antiulcer; osteopathic; dermatological; antiallergic; antidiabetic;
XX CC antidiabetic; cytostatic; neuroprotective; antidepressant; neurotropic;
XX CC antiparkinsonian; and immunostimulant. The proteins and polynucleotides
XX CC encoding them can be used in gene therapy, antisense therapy and vaccine
XX CC production. The proteins and polynucleotides are useful for screening for
XX CC agonists or antagonists of a protein and for the treatment and diagnosis
XX CC of disorders associated with the activity of a protein e.g. inflammation,
XX CC rheumatoid arthritis, septic shock, pancreatitis, cardiac dysfunction,
XX CC neuropathology, cardiac anaphylaxis, viral, bacterial, HIV and fungal
XX CC infections, autoimmunity, genetic diseases, haematopoietic disorders,
XX CC anaemia, platelet disorders, thrombocytopaenia, wounds, burns, ulcers,
XX CC osteoporosis, severe combined immunodeficiency, eczema, allergic
XX CC rhinitis, asthma, diabetes, cancer, multiple sclerosis, depression,
XX CC Alzheimer's disease, Parkinson's disease, neurodegenerative and
XX CC neurological disorders.

XX SQ Sequence 535 AA;

Alignment Scores:				Length:			
Pred. No.:	2.03e-39	584.00	Matches:	535			
Score:	45.00%	163	Conservative:	71			
Best Local Similarity:	31.35%	156	Mismatches:	130			
Query Match:	18.68%	19	Indels:	130			
DB:	22	22	Gaps:	19			
US-09-762-021A-1 (1-1710) x AAM25595 (1-535)							
Qy	131	GAGCAGGCACCATCTCTGGAGCCGGGATCCCTCCAGACAGCCCCACGAGGACCCCTA	190				
Db	10	GlulysIleArgGlnArgGlnSerIleLeuProPro	21				
Qy	191	GAGCAGACGCTCCACCATCCCAAGGCCCTGCCA---CGCCACACACAGTCCCGCAGAA	247				
Db	22	-----ProGlnGlyProAlaProIleProPheGlnHisArgGlyGlyAspSer	37				
Qy	248	CAAAGTCCC-----TTTACTCTGCCTCTCTCCCAAGCGG	280				
Db	38	ProGluAlaLysAsnArgValGlyProGlnValProLeuSerGluProGlyPheArgArg	57				
Qy	281	TCCTCTTCCCCTCCGAGGACCCA-----GAGAGGACGAGGAA	316				
Db	58	ArgGluSerGlnGluProArgAlaValLeuAlaGlnLysIleGluLysGluThrGln	77				
Qy	317	GTGCTGAACCATGTCCTAAGGACATTCAGCTGTTCATGGGAAGCTGGAGAAGGCC---	373				
Db	78	IleLeuAsnCysAlaLeuAspIleGluTrpPheValAlaArgLeuGlnLysAlaAla	97				
Qy	374	-----CAGGCAAGACACGAGGAGAAATTTGGGAAAAA	412				
Db	98	GluAlaPheLysGlnLeuAsnGlnArgLysLysGlyLysLysLys-----GlyLysLys	115				
Qy	413	AACAAGGACGAGGAGGTCTCAC-----CAGGCACAGTACATT	451				
Db	116	AlaProAlaGluGlyValLeuThrArgAlaArgProProSerGluGlyGluPheIle	135				
Qy	452	GACTGCTTCCAGAGATCAAGTACAGCTTCAACCTCTCTGGGAGGCTGGCCACCTGGCTG	511				
Db	136	AspCysPheGlnLysIleLysLeuAlaIleAsnLeuLeuAlaLysLeuGlnLysHisIle	155				
Qy	512	AAGGAGACAGTGCCTGAGCTGTACATCTCTTCAAGTCCCTCAAGTTCATCTCCTG	571				
Db	156	GlnAsnProSerAlaAlaGluLeuValHisPheLeuPheGlyProLeuAspLeuIleVal	175				
Qy	572	GCCAGGTCCCTGAGGCTGGCTTAGCAGCCCAAGTGATCTCACCCCTCTCCACCCCTAAA	631				
Db	176	AsnThrCysSerGlyProAspIleAlaArgSerValSerCysProLeuLeuSerArgAsp	195				
Qy	632	GCTATCAACCTGTACAGTCTGTCTAAGCCCACTGAGAGTAACTTTGGATGGGTGG	691				
Db	196	AlaValAspPheLeuArgGlyHisLeuValProLysGluMetSerLeuTrpGluSerLeu	215				
Qy	692	GGCCAGCCTGGACCACTAGCCGGGCGACGTGAGACGCGATGAGCCCTGCC---TAC	748				
Db	216	GlyGluSerTrpMetArgProArgSerGluTrpProArgGluProGlnValProLeuTyr	235				
Qy	749	CAACCCCATCTCAGATGACTGGCAACTTCCA-----	781				
Db	236	ValProLysPheHisSerGlyTrpGluProProValAspValLeuGlnAlaProTrp	255				
Qy	782	-----GAGCCCTCCAGCACACCTTAGGATACGAGGACCTGTTCCCTTCGCGG	835				
Db	256	GluValGluGlyLeuAlaSerAlaProIleGluValSerProValSer-----Arg	273				
Qy	836	GGAAGTCATAGTTAGGAGCACCTTCACACTTCTCAGGAGAGACACACACCATGAC	895				
Db	274	GlnSerIleArg---AsnSerGlnLysHisSerPro-----ThrSerGluProThr	289				
Qy	896	CCTCAGCTGGGAC-----CCCAACTCCAGGCCCTCCAGCCCAACCTGCCAG	946				

Db 290 Pro---ProGlyAspAlaLeuProProValSerSerProHisThrHisArgGlyTyrGln 308
QY 947 CCA-----GCCCTGAAATGCAAGTCTTCTAGCAGTTTGAAGCTAGGAACCCA 994
Db 309 ProThrProAlaMetAlaLysTyrValLysIleLeuTyrAspPheThrAlaArgAsnAla 328
QY 995 CGGGAAGTACTGCTGCTCCAGGAGAGAAGCTGGAGTTCTTGACACACAGCAAGCGGTGG 1054
Db 329 AsnGluLeuSerValLeuLysAspGluValLeuGluValLeuGluAspGlyArgGlnTrp 348
QY 1055 TGCTGCTGAAGATAGCGGGACGAGCGGCTGACTTCCAAAGCAACATCTCTG----- 1108
Db 349 TrpLysLeuArgSerArgSerGlyGlnAlaGlyTyrValProCysAsnIleLeuGlyGlu 368
QY 1108 ----- 1108
Db 369 AlaArgProGluAspAlaGlyAlaProPheGluGlnAlaGlyGlnLysTyrTrpGlyPro 388
QY 1109 -----CAGCCCTTACAGCCGGGACCCCTGGGACCCAGGGCCAG----- 1147
Db 389 AlaSerProThrHisLysLeuProProSerPheProGlyAsnLysAspGluLeuMetGln 408
QY 1148 -----TCACCC 1153
Db 409 HisMetAspGluValAsnAspGluLeuIleArgLysIleSerAsnIleArgAlaGlnPro 428
QY 1154 TCTCGG-----GTTCCAATG-----CTTCGACTTACGTCG 1183
Db 429 GlnArgHisPheArgValGluArgSerGlnProValSerGlnProLeuThrTyrGluSer 448
QY 1184 AGCCCTGAAGAGTCAACAGCTGGCTGCAGGACAGAGAACTTCTCCACTGCCACGGTGAGG 1243
Db 449 GlyProAspGluValArgAlaTrpLeuGluAlaLysAlaPheSerProArgIleValGlu 468
QY 1244 ACATTGGTCCCTGAGGGGACGAGCTACTTCGCATAGACCTGGGAGGTACAGATG 1303
Db 469 AsnLeuGlyIleLeuThrGlyProGlnLeuPheSerLeuAsnLysGluGluLeuTyrLys 488
QY 1304 CTATGTCCACAGAGGCCCCACCAATCTCTCCCGCTGGAGGCTCTCAGAGGATGCTG 1363
Db 489 ValCysGlyGluGluGlyValArgValTyrSerGlnLeuThrMetGlnLysAlaPheLeu 508
RESULT 12
ID AAM93529 standard; Protein; 723 AA.
XX AAM93529;
AC AAM93529;
DT 06-NOV-2001 (first entry)
XX Human polypeptide, SEQ ID NO: 3267.
DE Human; full length cDNA; cDNA synthesis; oligo-capping.
XX Homo sapiens.
OS Homo sapiens.
PN EP1130094-A2.
XX EP1130094-A2.
XX 05-SEP-2001.
PF 07-JUL-2000; 2000EP-0114089.
XX 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX (HELI-) HELIX RES INST.
XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.
DR N-PSDB; AAK94459.

XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX Claim 8; SEQ ID NO 3267; 1380pp + sequence listing; English.
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX SQ Sequence 723 AA:
Alignment Scores: 5.24e-39 Length: 723
Pred. No.: 579.50 Matches: 170
Score: 44.39% Conservative: 83
Best Local Similarity: 29.82% Mismatches: 184
Query Match: 18.53% Indels: 133
DB: 22 Gaps: 19
US-09-762-021a-1 (1-1710) x AAM93529 (1-723)
QY 2 GCAGAGCGACTGAAGACCAAGCGCTCGAAGAGCTCTGGAGGAGAGCTGGAGCAAGA--- 58
Db 144 AlaGluLeuIleArgGluAspIleGlnGlyAlaLeuHisAsnTyrArgSerGlyArgGly 163
QY 59 ---CCTCGACTTGGAGCCTTCAGCCAGCCAGGAC-----AGATGGAGGGGCGCT 106
Db 164 GluArgArgAlaAlaAlaLeuArgAlaThrGlnGluGluGlnArgAspArgSerPro 183
QY 107 GCTATGGAAGCGCGCTCCCTATGGAGCAGCAGCTATCTGGAGCGGGGATCCCTCCA 166
Db 184 AlaAlaGluThrPro---ProLeuGlnArgArg-----ProSer 195
QY 167 GAACAGCCCCCAGAGGACCTTAGACACAGCCTC-----CCACCATCCCCAAGG 217
Db 196 ValArgAlaValIleSerThrValGluArgGlyAlaGlyArgGlyArgProGlnAlaLys 215
QY 218 CCCCTGCCAGCCACACAGTGCCTGCGAGAACCAAGTGCCTTTACTCTGCTCTCCCAAG 277
Db 216 ProIleProGluAlaGluGluAlaGlnArgProGluProValGlyThrSerSerAsnAla 235
QY 278 CGGTCTCTTCCCGGAG-----GACCCAGAG 304
Db 236 AspSerAlaSerProAspLeuGlyProArgGlyProAspLeuAlaValLeuGlnAlaGlu 255
QY 305 AGGACGAGGAAGTGTGAACCATGTCTTAAGGACATTTCAGCTGTTCATGGGAAAGCTG 364
Db 256 ArgGluValAspIleLeuAsnHisValPheAspAspValGluSerPheValSerArgLeu 275
QY 365 GAGAGGCC-----CAGGCAAGACCCAGAGAAATAATTTGGGAAAAAACAAG 418
Db 276 GlnLysSerAlaGluAlaAlaArgValLeuGluHisArgGluArgGlyArgSerArg 295
QY 419 -----GACCAGGAGGTCTCAC-----CAGGCACAG 445
Db 296 ArgArgAlaAlaGlyGluGlyLeuLeuThrLeuArgAlaLysProProSerGlnAlaGlu 315
QY 446 TACATTGACTGCTTCCAGAGATCAAGTACAGCTTCAACCTCTCTGGGAAGGTGGCCACC 505
Db 316 TyrThrAspValLeuGlnLysIleLysTyrAlaPheSerLeuLeuAlaArgGly 335
QY 506 TGGCTGAAGAGACAAGTCCCTGAGCTGTCACATCCTTTCAGTCCCTTGAACCTTC 565
::: ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 22 -----ProGlnGlyProAlaProIleProSerSerThrAlaAlaGlyIlePro 37
Qy 242 CGAAGACCAAGT-----GCCTTTACTCTGCCTCCCTCCCAAGCGG 280
Db 38 ArgArgProArgIleAlaTrpAlaArgArgTrpProLeuSerGluProGlyPheArgarg 57
Qy 281 TCCTCTTCCCGCAGGACCCA-----GAGAGGAGCAGGAAA 316
Db 58 ArgGluSerGlnGluProArgAlaValLeuAlaGlnLysIleGluLysGluThrGln 77
Qy 317 GTGCTGAACCATGCTTAAGGACGATTCAGCTGTTCATGGGAAGCTGGGAAGGCC--- 373
Db 78 IleLeuAsnCysAlaLeuAspIleGluTrpPheValAlaAlaArgLeuGlnLysAlaAla 97
Qy 374 -----CAGCAAGACACGACGAGGAGAGAAATTTGGGAAGAAA 412
Db 98 GluAlaPheLysGlnLeuAsnGlnArgLysLysGlyLysLysLys-----GlyLysLys 115
Qy 413 AACAAAGGACACGAGGAGGTCTCACC-----CAGGCACAGTACATT 451
Db 116 AlaProAlaGluGlyValLeuThrLeuArgAlaArgProProGluGluGlyPheIle 135
Qy 452 GACTGCTTCCAGAGATCAAGTACAGCTTCAACCTCTCGGGAAGCTGGCCACCTGGCTG 511
Db 136 AspCysPheGlnLysIleLysLeuAlaIleAsnLeuLeuAlaLysLeuGlnLysHisIle 155
Qy 512 AAGGAGACAAGTGCCTGAGCTGTCACATCTCTTCAAGTCCCTGAATTCATCTG 571
Db 156 GlnAsnProGlnArgArgAspValValHisPheLeuPheGlyProLeuAspLeuIleVal 175
Qy 572 GCCAGGTGCTGAGCGTGGCTGACGACCCCAAGTGATCTCACCCCTCTCCACCCCTAAA 631
Db 176 AsnThrCysSerGlyProAspIleAlaArgSerValSerCysProLeuPheSerArgAsp 195
Qy 632 GCTATCAACCTGTACAGTCTCTGTCTAGCCGACCTGAGATTAACCTTTGATGGGGTGG 691
Db 196 AlaValAspPheLeuArgGlyHisLeuValProLysLeuMetSerLeuTrpGluSerLeu 215
Qy 692 GSCCCAGCTCGACCACTAGCGGCGGCGACCTGGACAGCGGATGACCCCTGCCC---TAC 748
Db 216 GlyGluSerTrpMetArgProArgSerGluTrpProArgGluProGlnValProLeuTrp 235
Qy 749 CAACCCACATTCTCAGATGACTGGCACTTCCTCA----- 781
Db 236 ValProLysPheHisSerGlyTrpGluProProValAspValLeuGlnGluAlaProTrp 255
Qy 782 -----GAGCCCTCCAGCAAGCACCTTAGATACAGGACCCCTGTTCCCTTCGCGGG 835
Db 256 GluValGluGlyLeuAlaSerAlaProIleGluGluValSerProValSer-----Arg 273
Qy 836 GGAAGTCATAGGTTAGGAGCACCTCACACTTCTCCTCAGGAGAACACACACCAATGAC 895
Db 274 GlnSerIleArg---AsnSerGlnLysHisArgProHisPheArgAlaHis----- 289
Qy 896 CCTCAGCTGGGAGCCCAAC-----TCCAGCGCCCTCCAGC 931
Db 290 ProProGlyGlyCysProTyHisGlnSerAlaProHisIleLeuThrArgGlyTrpGln 309
Qy 932 CCCAAACCTGCCCGCCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCTAGGAAC 991
Db 310 ProThrProAlaMet---AlaLysTrpValLysIleLeuTyAspPheThrAlaArgAsn 328
Qy 992 CCACGGGAACCTGACTGTGCTCCAGGAGAGAAAGCTGGAGGTTCTGGACCACAGCAAGCGG 1051
Db 329 AlaAsnGluLeuSerValLeuLysAspGluValLeuGluValLeuLysGlyArgGln 348
Qy 1052 TGGTGGCTGGTGAAGATCAGCGGCGGAGCGGCTACATTCCAAGCAACATCTCGAG 1111
Db 349 TrpTrpLysLeuArgSerArgSerGlyGlnAlaGlyTrpValProGlyAsnIleLeuGly 368
Qy 1112 CCCCTACAGCGC----- 1126
Db 369 GluAlaArgProGluAspAlaGlyAlaProPheGluGlnAlaGlyGlnLysTrpLeuGly 388

Qy 1127 ACCCTGGAGCCAGGCGCAGTCACTACCCTCTCGGTTCCTCA----- 1165
Db 389 ThrProGlnProAspProGlnAlaThrProLysLeuProGlyGlyThrIleAspGluLeu 408
Qy 1165 ----- 1165
Db 409 MetGlnHisMetAspGluValAsnAspGluLeuIleArgLysIleThrThrSerArgAla 428
Qy 1166 -----ATCCTTCGACTT 1177
Db 429 GlnProGlnArgHisPheArgValGluArgSerGlnProValSerGlnProLeuThrTrp 448
Qy 1178 AGCTCGAGGCTGAAGAGGTACAGACTGGCTGCAGGCAGAGAACTTCTCCACTGCCACG 1237
Db 449 GluSerGlyProAspGluValArgAlaPheLeuGluAlaLysAlaPheSerProArgIle 468
Qy 1238 GTGAGGACACTTGGTCCCTGACGGGAGCGACTACTTCCCATTAAGACCTGGGGAGCTA 1297
Db 469 ValGluAsnLeuGlyIleLeuThrGlyProGlnLeuPheSerLeuAsnLysGluGluLeu 488
Qy 1298 CAGATGCTATGTCACAGGAGGCCCCCAGCAATCTCTGCCGGCTGGAGGCTGTCTCAGAAG 1357
Db 489 LysLysValCysGlyGluGluGlyPheArgValTrpSerGlnLeuThrMetGlnLysAla 508
Qy 1358 ATGCTGGG-GATAAGCCCTTAGCCAGCAGCTTAGACACCTCCAAGAACCCGCGCTG 1416
Db 509 PheLeuGluLysGlnGlnSerGlySerGluLeuGluGluLeuMetAsnLysPheHisSer 528
Qy 1417 ATGCAAGATGCCAGATCTGATACCCATTTAGAGCCCCCGAGAATTCTCTCTTCTGGATCCCAG 1476
Db 529 MetAsnGlnArgArgGlyGluAspGlnLeuGlyProAlaAlaLeuGlyTrpGlyLeuArg 548
Qy 1477 TTTCAGCAACCCACACA 1494
Db 549 ArgGlySerProProThr 554
RESULT 14
ID AAR35451 standard; Protein: 821 AA.
XX AAR35451;
XX 17-DEC-2001 (updated)
XX 25-AUG-1993 (first entry)
XX Mouse eps8.
XX Epidermal growth factor receptor; EGFR-pathway substrate; eps;
XX tyrosine kinase receptor; TKR; SH2; SH3; mitogenesis.
XX Mus musculus.
XX USN7935311-N.
XX 01-APR-1993.
XX 25-AUG-1992; 92US-0935311.
XX 25-AUG-1992; 92US-0935311.
XX (USSH) US DEPT HEALTH & HUMAN SERVICE.
XX Di Flore PP, Fazliol F;
XX WPI; 1993-159477/19.
XX N-PSDB; AAQ40730.
XX Epidermal growth factor receptor substrate, eps 8 - used to
XX enhance mitogenic response of cells to epidermal growth factor
XX Disclosure; Page 30-37; 40pp; English.
XX

CC Eps8 is a novel EGFR substrate. The protein bears the
 CC characteristic signatures of TKR substrates including SH2 and
 CC SH3 domains. Eps8 is involved in the transduction of mitogenic
 CC signals and it can be used to enhance the mitogenic response of
 CC cells to EGF.
 CC (Note: Revised entry submitted to correct the patent number format of
 CC US Government-owned NTIS applications to prevent clashes with ongoing US
 CC granted patent numbers. For further information please visit the Derwent
 CC web site at www.derwent.com/dwpi/updates/ntis_us.html.)
 CC
 XX
 SQ Sequence 821 AA;

Alignment Scores: 2.91e-37 Length: 821
 Pred. No.: 558.50 Matches: 158
 Score: 558.50
 Percent Similarity: 38.82% Conservative: 78
 Best Local Similarity: 25.99% Mismatches: 171
 Query Match: 17.86% Indels: 201
 DB: 14 Gaps: 17

US-09-762-021a-1 (1-1710) x AAR35451 (1-821)

Qy 101 GGGCTGCTATGGAAGCCCTCCCTATGGAGCAGGACGCTATCTGGAGCGGGGATC 160
 Db 190 GlyLysGlnTysArgArgProGluAlaLeuArgMetIleAlaLysAlaAspProGlyIle 209
 Qy 161 CTTCCAGAACGCCACCAGAGGACCTAGAGCAGACGCTCCACCACCTCCCAAGGCC 220
 Db 210 ProProPro-----ProArgAlaProAlaPro 218
 Qy 221 CTGCCAGCCACACC-----AGTGCC 241
 Db 219 ValProGlyThrValThrGlnValAspValArgSerArgValAlaAlaTyrSerAla 238
 Qy 242 CGAAGAACCAAGTCTTACTCTCCCTCCCAAGGGGCTCTCTCCCGGAGAC---- 298
 Db 239 TrpAlaAlaAspGlnGlyAspPheGluLysProArgGlnTyrHisGluGlnGluThr 258
 Qy 299 CCAGAG-----AGGACGAGGAAGTGTGAACCATGTCTCTAAGG 337
 Db 259 ProGluMetMetAlaAlaArgIleAspArgValGlnIleLeuAsnHisIleLeuAsp 278
 Qy 338 GACATTGAGCTGTTTCATGGGAAGCTGGAGAGGCCAG-----GCA 379
 Db 279 AspIleGluPheIleThrLysLeuGlnLysAlaAlaGluAlaPheSerGluLeuSer 298
 Qy 380 AAGACCAGCAGAGAGAAATTTGGGAAAAAACAAGGACAGGAGGTCTCACC---- 436
 Db 299 LysArgLysLysSerLysLysSerLysArgLysGlyProGlyGluGlyValLeuThrLeu 318
 Qy 437 -----CAGGCACAGTACATTGACTGCTTCCAGAAAGATCAAGTACAGC 478
 Db 319 ArgAlaLysProProProProAspGluPheValAspCysPheGlnLysPheLysHisGly 338
 Qy 479 TTCACCTCTCGGAAGGCTGGCCACCTGGCTGAAGAGAGACAAGTCCCTGAGCTGTA 538
 Db 339 PheAsnLeuLeuAlaLysLeuLysSerHisIleGlnAsnProSerAlaSerAspLeuVal 358
 Qy 539 CACATCTCTTCAAGTCCCTCAACTTCTCTGCGCAGGTCCTCAGGCTGGCCTAGCA 598
 Db 359 HisPheLeuPheThrProLeuAsnMetValValGlnAlaThrGlyGlyProGluLeuAla 378
 Qy 599 GCCCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGCTACAGTCTGTCTA 658
 Db 379 SerSerValLeuProProLeuLeuThrLysAspThrValAspPheLeuAsnTyrThrAla 398
 Qy 659 AGCCACCTGAGATAACCTTTGGATGGGTGGGCCCGCAGCTGGACCACTAGCGGGGCC 718
 Db 399 ThrAlaGluGluArgLysLeuTyrMetSerLeuGlyAspSerTrpValLysValArgAla 418
 Qy 719 GACTGGACAGCGATGAGCCCTG-----CCCTACCAACCCACATTCTCAGATGAGTGGCAA 775
 Db 419 GluTrpProLysGluGlnPheIleProProTyrValProArgPheArgAsnGlyTrpGlu 438

Qy 776 CTTCCTCA-----GAGCCCTCCAGCAAGCACCTTAGGTATACAGCAGC 817
 Db 439 ProProMetLeuAsnPheMetGlyAlaProThrGluGlnAspMet-----TyrGlnLeu 456
 Qy 818 CCTGTTTCCCTTCGCGGGGAAGTCAT---AGTTAGGAGCAGCCTCACACTTCTCTCAG 874
 Db 457 AlaGluSerValAlaAlaAsnAlaGluHisGlnArgLysGlnAspSerLysArgLeuSerThr 476
 Qy 875 GAGAAGACACACACCATGACCTCAGCTCGGGGAC----- 910
 Db 477 GluHisSerAsnValSerAspTyrProProAlaAspGlyTyrAlaTyrSerSerSerMet 496
 Qy 911 -----CCC 913
 Db 497 TyrHisArgGlyProHisAlaAspHisGlyGluAlaAlaMetProPheLysSerThrPro 516
 Qy 914 AAC-----TCCAGGCGCTCCAGCCGCCAACCTCCAGCAGCAGCCCTGAAATGCAA 964
 Db 517 AsnHisGlnValAspArgAsnTyrAspAlaValLysThrGlnProLysLysTyrAlaLys 536
 Qy 965 GTCCTGTACAGTTTGAAGCTAGTAACCCAGCGGAAGTCTGTGTCAGGAGAGAG 1024
 Db 537 SerLysTyrAspPheValAlaAlaArgAsnSerGluLeuSerValMetLysAspVal 556
 Qy 1025 CTGAGGTTCTGGACACAGCAACGGTGTGTGTGAAGATGAGCGGGGAGC 1084
 Db 557 LeuGluLeuAspArgArgGlnTrpLysValArgAsnAlaSerGlyAspSer 576
 Qy 1085 GGCTACATTCCAAGCAATCCTGGAG----- 1111
 Db 577 GlyPheValProAsnAsnIleLeuAspIleMetArgThrProGluSerGlyValGlyArg 596
 Qy 1111 ----- 1111
 Db 597 AlaAspProProTyrThrHisThrIleGlnLysGlnArgThrGluTyrGlyLeuArgSer 616
 Qy 1112 -----CCCTCA 1117
 Db 617 AlaAspThrProSerAlaProSerProProThrProAlaProValProValProLeu 636
 Qy 1118 CAGCGGGGACCTCTGG----- 1135
 Db 637 ProProSerValProAlaProValSerValProLysValProAlaAspValThrArgGln 656
 Qy 1135 ----- 1135
 Db 657 AsnSerSerSerSerSerGlyGlySerIleValArgAspSerGlnArgTyrLysGln 676
 Qy 1135 ----- 1135
 Db 677 LeuProValAspArgArgLysSerGlnMetGluGluValGlnAspGluLeuPheGlnArg 696
 Qy 1136 ---ACCCAGGGCCAGTCACCTCTCGG-----GTTCCA 1165
 Db 697 LeuThrIleGlyArgSerAlaAlaGlnArgLysPheHisValProArgGlnAsnValPro 716
 Qy 1166 ATGCTTCGACTTAC-----TCAGAGCCCTGAAGAGGTTCACAGACTGGCTGCAGCAGC 1219
 Db 717 ValIleAsnIleThrTyrAspSerSerProGluGluValLysThrTrpLeuGlnSerLys 736
 Qy 1220 AACTTCTCCACTGCGCAGGTGAGCACACTTGGTCCCTGAGCGGGAGCAGCTACTTCGC 1279
 Db 737 GlyPheAsnProValThrValAsnSerLeuGlyValLeuAsnGlyAlaGlnLeuPheSer 756
 Qy 1280 ATAAGACCTGGGAGCTACAGATGCTATGTCCACAGGAGGCCGCCAGCAANTCTCTCCCG 1339
 Db 757 LeuAsnLysAspGluLeuArgSerValCysPro---GluGlyAlaArgValPheAsnGln 775
 Qy 1340 CTGAGGCTGTCAAGAGGATGCTG 1363
 Db 776 IleThrValGlnLysAlaAlaLeu 783

RESULT 15
AAG75249
ID AAG75249 standard; Protein; 112 AA.
XX
AC AAG75249;
XX
DT 03-SEP-2001 (first entry)
XX
Human colon cancer antigen protein SEQ ID NO:6013.
XX
DE Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma.
XX
OS Homo sapiens.
XX
PN WO200122920-A2.
XX
PD 05-APR-2001.
XX
PF 28-SEP-2000; 2000WO-US26524.
XX
PR 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX
DR WPI: 2001-235357/24.
DR N-PSDB; AAH34654.
XX
PT Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX
PS Claim 11; Page 7506-7508; 9803pp; English.
XX
CC AAH32943 to AAH37195 and AAG773514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 112 AA;

Alignment Scores:
Pred. No.: 2,98e-31 Length: 112
Score: 482.00 Matches: 94
Percent Similarity: 96.00% Conservative: 2
Best Local Similarity: 94.00% Mismatches: 4
Query Match: 15.41% Indels: 0
DB: 22 Gaps: 0

US-09-762-021A-1 (1-1710) x AAG75249 (1-112)

QY 311 GAGGAAGTCTGAACCATGCTCTCAAGGACATTGAGCTGTTCAAGGAAAGCTGGAGAAG 370
Db 13 GInGluValLeuAsnHisValLeuArgAspIleGluLeuPheMetClyLysLeuGluLys 32
QY 371 GCCCAGGCAAGACCAGCAGGAAGAAGAAATTTGGGAAAAAACAAGGACCGAGGT 430
* * *

Db 33 AlaGlnAlaLysThrSer**LysLysLysPheGlyLysAsnLysAspGlnGlyGly 52
QY 431 CTCACCCAGGCACAGTACATTGCTTCCAGAAAGATCAAGTACAGCTTCAACCTCCTG 490
Db 53 LeuThrGlnAlaGlnTyrIleAspCysPheGlnLysIleLysHisSerPheAsnLeuLeu 72
QY 491 GGAAGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCTGAGCTGCTACACATCCTCTTC 550
Db 73 GlyArgLeuAlaThrTrpLeuLysGluThrSerAlaProGluLeuValHisIleLeuPhe 92
QY 551 AAGTCCCTGAACCTTCATCCTGGCCAGGTGCCCTGAGGTGGCCTAGCAGCCCAAGTGATC 610
Db 93 Lys***LeuAsnPhe***LeuAlaArgCysProGluAlaGly***AlaAlaGlnValIle 112

Search completed: February 25, 2003, 19:39:48
Job time : 53 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 17:06:18 ; Search time 17.5 Seconds

(without alignments)

5750.080 Million cell updates/sec

Title: US-09-762-021a-1

Perfect score: 3127

Sequence: 1 ggcagagcagctgaagacca.....aaaaaaaaaataangataaa 1710

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame_n2p.model -DEV=xlh
-O=/cgn2_1/USPTO.spool/US09762021/runat_24022003_153039_7605/app_query.fasta_1.1863
-DB-Issued_Patents_AA -QFMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09762021.ecgn_1.1.9 @runat_24022003_153039_7605 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/PTCUS_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	558.5	17.9	821	1	US-07-935-311A-4
2	558.5	17.9	821	1	US-08-368-079-4
3	558.5	17.9	821	5	PCT-US93-07996-4
4	200.5	6.4	1057	3	US-08-931-820-4
5	191	6.1	1078	3	US-08-963-825-21
6	191	6.1	1078	4	US-09-500-811-21
7	191	6.1	1078	4	US-09-570-573-21
8	191	6.1	1078	4	US-09-548-608-21
9	181.5	5.8	561	1	US-08-842-255-52
10	179.5	5.7	504	4	US-09-219-849-3
11	178	5.7	552	4	US-09-219-849-7
12	177.5	5.7	1461	4	US-09-585-887-9

13	177.5	5.7	1461	4	US-09-289-578-9	Sequence 9, Appli
14	176.5	5.6	633	1	US-08-642-255-73	Sequence 73, Appli
15	176.5	5.6	720	4	US-09-219-849-4	Sequence 4, Appli
16	176.5	5.6	777	1	US-08-642-255-53	Sequence 53, Appli
17	176.5	5.6	960	4	US-09-219-849-5	Sequence 5, Appli
18	175.5	5.6	1341	3	US-08-963-825-18	Sequence 18, Appli
19	175.5	5.6	1341	4	US-09-500-811-18	Sequence 18, Appli
20	175.5	5.6	1341	4	US-09-570-573-18	Sequence 18, Appli
21	175.5	5.6	1341	4	US-09-548-608-18	Sequence 18, Appli
22	174.5	5.6	1057	3	US-08-931-820-1	Sequence 1, Appli
23	172.5	5.5	1185	4	US-09-041-886-23	Sequence 23, Appli
24	171.5	5.5	960	4	US-09-219-849-6	Sequence 6, Appli
25	171.5	5.5	1274	4	US-09-095-443-2	Sequence 2, Appli
26	169	5.5	720	4	US-09-219-849-4	Sequence 4, Appli
27	169	5.5	777	1	US-08-642-255-53	Sequence 53, Appli
28	168.5	5.4	1065	1	US-08-642-255-72	Sequence 72, Appli
29	167	5.4	504	4	US-09-219-849-3	Sequence 3, Appli
30	167	5.4	561	1	US-08-642-255-52	Sequence 52, Appli
31	166	5.3	1060	3	US-08-931-820-3	Sequence 3, Appli
32	166	5.3	1418	3	US-08-963-825-20	Sequence 20, Appli
33	166	5.3	1418	4	US-09-010-999-1	Sequence 1, Appli
34	166	5.3	1418	4	US-09-500-811-20	Sequence 20, Appli
35	166	5.3	1418	4	US-09-570-573-20	Sequence 20, Appli
36	163.5	5.2	1442	2	US-09-548-608-20	Sequence 20, Appli
37	163.5	5.2	1442	2	US-08-316-650-12	Sequence 12, Appli
38	163.5	5.2	1442	5	PCT-US95-02251-12	Sequence 12, Appli
39	162	5.2	822	4	US-09-219-849-49	Sequence 49, Appli
40	158	5.1	1054	1	US-08-642-255-62	Sequence 62, Appli
41	157	5.0	357	1	US-07-609-716-66	Sequence 66, Appli
42	157	5.0	357	1	US-08-642-255-33	Sequence 33, Appli
43	157	5.0	357	4	US-08-475-411A-66	Sequence 66, Appli
44	157	5.0	357	4	US-08-478-029A-66	Sequence 66, Appli
45	155.5	5.0	1418	4	US-09-010-999-1	Sequence 1, Appli

ALIGNMENTS

RESULT 1

US-07-935-311A-4
; Sequence 4, Application US/07935311A
; Patent No. 5378809
; GENERAL INFORMATION:
; APPLICANT: Di Fiore, Pier Paolo
; APPLICANT: Fazioli, Francesca
; TITLE OF INVENTION: Substrate of the Epidermal Growth
; TITLE OF INVENTION: Factor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/935,311A
; FILING DATE: 19920825
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Israelsen, Ned A.
; REGISTRATION NUMBER: 29,655
; REFERENCE/DOCKET NUMBER: NIH035.001A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 821 amino acids

; TYPE: AMINO ACID
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-07-935-311A-4

Alignment Scores:

Pred. No.: 4,16e-39 Length: 821
 Score: 558.50 Matches: 158
 Percent Similarity: 38.82% Conservatives: 78
 Best Local Similarity: 25.99% Mismatches: 171
 Query Match: 17.86% Indels: 201
 DB: 1 Gaps: 17

US-09-762-021A-1 (1-1710) x US-07-935-311A-4 (1-821)

```

QY 101 GGCCTGTATGAAAGCGCGTCCCTATGGAGCAGCGCATCTATCGAGCGGGGATC 160
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 190 GlyLysGlnLysArgArgProGluAlaLeuArgMetIleAlaLysAlaAspProGlyIle 209
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 161 CTTCCAGAACAGCCACCAGAGGACCTTAGACACAGCCTCCCAACATCCCAAGGCC 220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 210 ProProPro-----ProArgAlaProAlaPro 218
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 221 CTGCCAGCCACACC-----AGTGCC 241
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 219 ValProProGlyThrValThrGlnValAspValArgSerArgValAlaAlaAlaTrpSerAla 238
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 242 CGAGAACCAAGTGCCTTACTCTGCTCTCCCAAGCGGTCTCTCTCCCGAGGAC--- 298
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 239 TrpAlaAlaAspGlnGlyAspPheGluLysProArgGlnTyrHisGluGlnGluThr 258
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 299 CGAGAG-----AGGGACGAGGAGTGTCTGAACCATCTCTCTTAAGG 337
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 259 ProGluMetMetAlaAlaArgIleAspArgAspValGlnIleLeuAsnHisIleLeuAsp 278
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 338 GACATTGACCTTTCATGGGAAGCTGGGTGAAGGACAAAGTGCCTGAGCTCGTA 379
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279 AspIleGluPhePheIleThrLysLysLeuGlnLysAlaAlaGluAlaPheSerGluLeuSer 298
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 380 AAGACCAGCAGCAGAGAAATTTGGGAAAAAACAAGGACCGGAGGTCTCACC--- 436
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 299 LysArgLysLysSerLysArgLysGlyProGlyGluGlyValLeuThrLeu 318
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 437 -----CAGCACAGTACATTGACTGCTCCAGAAAGTCAAGTACAGC 478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 319 ArgAlaLysProProProAspGluPheValAspCysPheGlnLysPheLysHisGly 338
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 479 TTCACCTCTCGGAGGCTGCCACCTGCTGGAAGGACAAAGTGCCTGAGCTCGTA 538
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 339 PheAsnLeuAlaLysLysLysSerHisIleGlnAsnProSerAlaSerAspLeuVal 358
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 539 CACATCTCTTCAAGTCCCTGAACCTTCATCTGCGCCAGGTGCGCTGAGGCTGCCGTAGCA 598
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 359 HisPheLeuPheThrProLeuAsnMetValValGlnAlaThrGlyGlyProGluLeuAla 378
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 599 GCCCAAGTGATCTACCCCTCTCACCCTTAAAGTATCAACCTGCTACAGTCTGTCTA 658
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 379 SerSerValLeuSerProLeuLeuThrLysAspThrValAspPheLeuAsnTyrThrAla 398
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 659 AGCCACCTGAGAGTAACCTTTGGATGGGTGGGCCAGCCCTGGACCACTAGCCGGGCC 718
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 399 ThrAlaGluGluArgLysLeuTrpMetSerLeuGlyAspSerTrpValLysValArgAla 418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 719 GACTGTGACAGCGGATGAGCCCTG---CCCTACCAACCCACATTTCTCAGATGACGTGCAA 775
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 419 GluTrpProLysGlnPheIleProProTyrValProArgPheArgAsnGlyTrpGlu 438
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 776 CTTCCA-----GAGCCTCCAGCCAGCACCCTTAGATACCAGGAC 817
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 439 ProProMetLeuAsnPheMetGlyAlaProThrGluGlnAspMet-----TyrGlnLeu 456
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 818 CCTGTTTCCCTTCGGGGGAGTCAAT---AGTTAGGAGCACCTCACACTTCTCCTCAG 874
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
  
```

RESULT 2

US-08-368-079-4
 ; Sequence 4, Application US/08368079
 ; Patent No. 5610018
 ; GENERAL INFORMATION:
 ; APPLICANT: DI Fiore, pier Paolo
 ; APPLICANT: Fazioli, Francesca

```

Db 457 AlaGluSerValAlaAsnAlaGluHisGlnArgLysGlnAspSerLysArgLeuSerThr 476
QY 875 GAGAGACACACACCATGACCTCAGCTGGGAC----- 910
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 477 GluHisSerAsnValSerAspTyrProProAlaAspGlyTyrAlaTyrSerSerSerMet 496
QY 911 -----CCC 913
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 497 TyrHisArgGlyProHisAlaAspHisGlyGluAlaAlaMetProPheLysSerThrPro 516
QY 914 AAC-----TCCAGGCCCTCCAGCCCCAAACCTGCCAGCCAGCCCTGAAAATGCAA 964
    ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 517 AsnHisGlnValAspArgAsnTyrAspAlaValLysThrGlnProLysLysTyrAlaLys 536
QY 965 GTCCTGTACGAGTTTCAAGCTAGGAAACCCAGCGGAACCTGACTGTGTCCAGGAGAGAAG 1024
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 537 SerLysTyrAspPheValAlaArgAsnSerSerGluLeuSerValMetLysAspVal 556
QY 1025 CTGGAGGTTCTGGACACACAGCAAGCGGTGGTGGTGAAGTGAAGTGAAGTGAAGTGAAG 1084
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 557 LeuGluIleLeuAspArgArgGlnTrpTrpLysValArgAsnAlaSerGlyAspSer 576
QY 1085 GCCTACATTCACAGCAACATCTCTGAG----- 1111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 577 GlyPheValProAsnAsnIleLeuAspIleMetArgThrProGluSerGlyValGlyArg 596
QY 1111 ----- 1111
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 597 AlaAspProProTyrThrHisThrIleGlnLysGlnArgThrGluTyrGlyLeuArgSer 616
QY 1112 -----CCCTTA 1117
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 617 AlaAspThrProSerAlaProSerProProThrProAlaProAlaProValProValProLeu 636
QY 1118 CAGCCGGGAGCCCTCTGG----- 1135
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 637 ProProSerValProAlaProValSerValProLysValProAlaAspValThrArgGln 656
QY 1135 ----- 1135
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 657 AsnSerSerSerSerSerGlySerIleValArgAspSerGlnArgTyrLysGln 676
QY 1135 ----- 1135
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 677 LeuProValAspArgArgLysSerGlnMetGluGluValGlnAspGluLeuPheGlnArg 696
QY 1136 ---ACCCAGGCCAGTCACTCTCTCG-----GTTCCA 1165
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 697 LeuThrIleGlyArgSerAlaAlaGlnArgLysPheHisValProArgGlnAsnValPro 716
QY 1166 ATGCTTCGACTTAGC-----TCGAGGCTGAAGAGGTACACAGACTGCTGCAGGCAGAG 1219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 717 ValIleAsnIleThrTyrAspSerSerProGluGluValLysThrTrpLeuGlnSerLys 736
QY 1220 AACTTCTCCTGCTCCAGCGGTGAGGACACTGGTCCCTCCAGCGGAGGAGGAGGAGGAGGAG 1279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 737 GlyPheAsnProValThrValAsnSerLeuGlyValLeuAsnGlyAlaGlnLeuPheSer 756
QY 1280 ATAAGACCTGGGAGCTACAGATGCTATGTCACAGAGGAGGAGGAGGAGGAGGAGGAGGAG 1339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 757 LeuAsnLysAspGluLeuArgSerValCysPro---GluGlyAlaArgValPheAsnGln 775
QY 1340 CTGGAGGCTGTGAGAGGATGCTG 1363
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 776 IleThrValGlnLysAlaAlaLeu 783
  
```

TITLE OF INVENTION: eps8, A Substrate for the Epidermal Growth Factor
TITLE OF INVENTION: Kinase, Antibodies Thereto, and Methods of Use Thereof

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSEE: Knobbe, Martens, Olson & Bear

STREET: 620 Newport Center Drive, 16th Floor

CITY: Newport Beach

STATE: CA

COUNTRY: USA

ZIP: 92660

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq Version 1.5

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/368,079

FILING DATE: 03-JAN-1995

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/935,311

ATTORNEY/AGENT INFORMATION:

NAME: Israel, Ned A.

REGISTRATION NUMBER: 29,655

REFERENCE/DOCKET NUMBER: NIH035.001DV1

TELEPHONE: (619) 235-8550

TELEFAX: (619) 235-0176

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 821 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-368-079-4

Alignment Scores:

Pred. No.:	4.16e-39	Length:	821
Score:	558.50	Matches:	158
Percent Similarity:	38.82%	Conservative:	78
Best Local Similarity:	25.99%	Mismatches:	171
Query Match:	17.86%	Indels:	201
DB:	1	Gaps:	17

US-09-762-021a-1 (1-1710) x US-08-368-079-4 (1-821)

Qy	101	GGGCTGCTATGGAAAGCCGCTCCCTATGGAGCAGGCAGCTATCTGGAGCCGGGATC	160
Db	190	GlyLysGlnLysArgArgProGluAlaLeuArgMetIleAlaLysAlaAspProGlyIle	209
Qy	161	CTCCAGAACACCCACCAGAGCCCTAGAGCACGCCCTCCACCATCCCCAGGCCC	220
Db	210	ProProPro-----ProArgAlaProAlaPro	218
Qy	221	CTGCCACCCACAC-----AGTGCC	241
Db	219	ValProProGlyThrValThrGlnValAspValArgSerArgValAlaAlaTrpSerAla	238
Qy	242	CGAGAACCAAGTGCTTACTGCTCTCCCAAGCGGCTCTCTCCCGGAGAC---	298
Db	239	TrpAlaAlaAspGlnGlyAspPheGluLysProArgGlnTyrHisGluGlnGluLuphr	258
Qy	299	CCAGAG-----AGGACGAGGAGAGTCTGACCATGCTCTAAGG	337
Db	259	ProGluMetMetAlaAlaArgIleAspArgValGlnIleLeuAsnHisIleLeuAsp	278
Qy	338	GACATTGAGCTTCTCGGAAAGCTGGAGAGGCCAG-----GCA	379
Db	279	AspIleGluPhePheIleThrLysLeuGlnLysAlaAlaGluAlaPheSerGluLeuSer	298
Qy	380	AAGACACGAGGAAGAAATTTGGGAAAAAACAGGACCGAGGAGGTCTCACC---	436

Db	299	LysArgLysLysSerLysLysSerLysArgLysGlyProGlyGluGlyValLeuThrLeu	318
Qy	437	-----CAGGCACAGTACATTGACTCTTCCAGAACATCAAGTACAGC	478
Db	319	ArgAlaLysProProProAspGluPheValAspCysPheGlnLysPheLysHisGly	338
Qy	479	TTCACCTCTCGGAAGCTGGCCACCTGGCTGAAGGAGACAAAGTCCCTCAGCTCGTA	538
Db	339	PheAsnLeuLeuAlaLysLysSerHisIleGlnAsnProSerAlaSerAspLeuVal	358
Qy	539	CACATCTCTTCAAGTCCCTGAACCTTCTCTGCGCAGCTGCCCTGAGGCTGGCCTACA	598
Db	359	HisPheLeuPheThrProLeuAsnMetValValGlnAlaThrGlyGlyProGluLeuAla	378
Qy	599	GCCCAAGTATCTCACCCCTCTCACCCCTAAAGTATCAACCTGCCTACAGTCTGTCTA	658
Db	379	SerSerValLeuSerProLeuLeuThrLysAspThrValAspPheLeuAsnTyrThrAla	398
Qy	659	AGCCACCTGAGAGTAACTTTTGGATGGGTTGGCCCGCAGCTGGACCACTAGCCGGCC	718
Db	399	ThrAlaGluGluArgLysLeuTrpMetSerLeuGlyAspSerTrpValLysValArgAla	418
Qy	719	GACTGGACAGCGGATGAGCCCTG---CCCTACCAACCCACATCTCTCAGATGACTGCAA	775
Db	419	GluTrpProLysGluGlnPheIleProTrpValProArgPheArgAsnGlyTrpGlu	438
Qy	776	CTTCCA-----GAGCCCTCCAGCCAGCACCTTAGGATACCAAGC	817
Db	439	ProProMetLeuAsnPheMetGlyAlaProThrGluGlnAspMet-----TyrGlnLeu	456
Qy	818	CCTGTTTCTCTCGCGGGGAAGTCAT---AGGTTAGGAGCAGCCTCACACTTCTCTCAG	874
Db	457	AlaGluSerValAlaAsnAlaGluHisGlnArgLysGlnAspSerLysArgLeuSerThr	476
Qy	875	GAGAAGACACACACCATGACCTCAGCCTGGGAC-----	910
Db	477	GluHisSerAsnValSerAspTyrProProAlaAspGlyTyrAlaTyrSerSerMet	496
Qy	911	-----CCC	913
Db	497	TyrHisArgGlyProHisAlaAspHisGlyGluAlaAlaMetProPheLysSerThrPro	516
Qy	914	AAC-----TCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGAAAAATGCAA	964
Db	517	AsnHisGlnValAspArgAsnTyrAspAlaValLysThrGlnProLysLysTyrAlaLys	536
Qy	965	GTCCTGTACAGTGTGAAGCTAGGAACCCAGCGGAACCTGACTGTGTCCAGGAGAGAG	1024
Db	537	SerLysTyrAspPheValAlaArgAsnSerSerGluLeuSerValMetLysAspVal	556
Qy	1025	CTGAGGTTCTGGACACAGCGGTGGTGGTGGTGAAGATGAGCGCGGAGGAGC	1084
Db	557	LeuGluIleLeuAspArgArgGlnTrpLysValArgAsnAlaSerGlyAspSer	576
Qy	1085	GGCTACATTCCAAGCAACATCTGGAG-----	1111
Db	577	GlyPheValProAsnAsnIleLeuAspIleMetArgThrProGluSerGlyValGlyArg	596
Qy	1111	-----	1111
Db	597	AlaAspProProTyrThrHisThrIleGlnLysGlnArgThrGluTyrGlyLeuArgSer	616
Qy	1112	-----CCCCCTA	1117
Db	617	AlaAspThrProSerAlaProSerProProThrProAlaProValProValProLeu	636
Qy	1118	CAGCCGGGACCCCTGGG-----	1135
Db	637	ProProSerValProAlaProValSerValProLysValProAlaAspValThrArgGln	656
Qy	1135	-----	1135
Db	657	AsnSerSerSerAspSerGlyGlySerIleValArgAspSerGlnArgTyrLysGln	676

QY 1135 ----- 1135
Db 677 LeuProValAspArgLysSerGlnMetGluGluValGlnAspGluLeuPheGlnArg 696
QY 1136 ----ACCCAGGGCAGTCACCTCTCGG-----GTTCCA 1165
Db 697 LeuThrIleGlyArgSerAlaAlaGlnArgLysPheHisValProArgGlnAsnValPro 716
QY 1166 ATGCTTCGACTTACG-----TCGAGGCTGAAGAGTCACAGCTGGCTGCAGGCAGAG 1219
Db 717 ValIleAsnIleThrTyrAspSerProGluGluValIleThrTrpLeuGlnSerLys 736
QY 1220 AACTTCTCCACTGCCAGCTGAGGACATTGGTCTCCTGACGGGAGCCAGCTACTTTCGC 1279
Db 737 GlyPheAsnProValThrValAsnSerLeuGlyValLeuAsnGlnYalaGlnLeuPheSer 756
QY 1280 ATAAGACCTGGGAGCTACAGATGCTATGCTCCACAGGAGGCCCGCCAGCATCTGTCCCGG 1339
Db 757 LeuAsnLysAspGluLeuArgSerValCysPro---GluGlyAlaArgValPheAsnGln 775
QY 1340 CTGGAGGCTGTGACAGAGGATGCTG 1363
Db 776 IleThrValGlnLysAlaAlaLeu 783

RESULT 3

PCT-US93-07996-4
; Sequence 4. Application PC/TUS9307996
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States, as represented by the
; APPLICANT: Secretary of Health and Human Services
; TITLE OF INVENTION: Substrate of the Epidermal Growth Factor Kinase
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Knobbe, Martens, Olson & Bear
; STREET: 620 Newport Center Drive, 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/07996
; FILING DATE: 19930825
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 235-8550
; TELEFAX: (619) 235-0176
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 821 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
PCT-US93-07996-4

Alignment Scores:
Pred. No.: 4.16e-39 Length: 821
Score: 558.50 Matches: 158
Percent Similarity: 38.82% Conservative: 78
Best Local Similarity: 25.98% Mismatches: 171
Query Match: 17.86% Indels: 201
DB: 5 Gaps: 17

US-09-762-021A-1 (1-1710) x PCT-US93-07996-4 (1-821)

QY 101 GGGCTGCTATGAAGGCGGCTCCCTATGGAGCGGACGCTATCTGGAGCGGGGATC 160
Db 190 GlyLysGlnLysArgArgProGluAlaLeuArgMetIleAlaLysAlaAspProGlyIle 209

QY 161 CCTCCAGAACAGCCCCACAGAGACCCTTAGACACAGCCTCCACCATCCCAAGGCC 220
Db 210 ProProPro-----ProArgAlaProAlaPro 218
QY 221 CTGCCACGCCACACC-----AGTGGC 241
Db 219 ValProProGlyThrValThrGlnValAspValArgSerArgValAlaAlaAlaTrpSerAla 238
QY 242 CGAAGAACCAAGTGCCTTACTCTGCTCCTCAAGCGGTCTCTTCCGCCCGAGGAC--- 298
Db 239 TrpAlaAlaAspGlnGlyAspPheGluLysProArgGlnTyrHisGluGlnGluGluThr 258
QY 299 CCAGAG-----AGGAGCAGAGAAAGTCTGAACCATGCTCTTAAG 337
Db 259 ProGluMetMetAlaAlaArgIleAspArgAspValGlnIleLeuAsnHisIleLeuAsp 278
QY 338 GACATTGACCTTTCATGGGAAAGCTGGAGAGGCCAG-----GCA 379
Db 279 AspIleGluPhePheIleThrLysLeuGlnLysAlaAlaGluAlaPheSerGluLeuSer 298
QY 380 AAGACCAGCAGGAAGAATAATTGGGAAAAAACAAGGACGAGGAGTCTCACCC--- 436
Db 299 LysArgLysLysSerLysSerLysArgLysGlyProGlyGluGlyValLeuThrLeu 318
QY 437 -----CAGGCACACAGTACATGTGCTTCCAGAAAGATCAAGTACAGC 478
Db 319 ArgAlaLysProProProProAspGluPheValAspCysPheGlnLysPheLysHisGly 338
QY 479 TTCACCTCTCTGGGAGGCTGGCCACCTGGCTGAAGGACACAGTCCCTCAGCTCGTA 538
Db 339 PheAsnLeuLeuAlaLysLysSerHisIleGlnAsnProSerAlaSerAspLeuVal 358
QY 539 CACATCTCTTCAAGTCCCTGAACCTTCATCTCGGCGAGGTGCCCTGAGGCTGGCTAGCA 598
Db 359 HisPheLeuPheThrProLeuAsnMetValValGlnAlaThrGlyClyProGluLeuAla 378
QY 599 GCCCAAGTGATCTACCCCTCTCCTACCCCTAAAGCTATCAACCTGCTACAGCTCTGCTCA 658
Db 379 SerSerValLeuSerProLeuLeuThrLysAspThrValAspPheLeuAsnTyrThrAla 398
QY 659 AGCCACCTGAGAGTAACCTTTGGATGGGTTGGGCGCCAGCTGGACACTACCGCGGCC 718
Db 399 ThrAlaGluGluArgLysLeuTrpMetSerLeuGlyAspSerTrpValLysValArgAla 418
QY 719 GACTGGACAGCGGATGAGCCCTG---CCCTACCAACCCACATCTCAGACTCTGCGCAA 775
Db 419 GluTrpProLysGluGlnPheIleProTyrValProArgPheArgAsnGlyTrpGlu 438
QY 776 CTCCA-----GAGCCCTCCAGCAGCAAGCACCCTTAGATACAGGAC 817
Db 439 ProProMetLeuAsnPheMetGlyAlaProThrGluGlnAspMet-----TyrGlnLeu 456
QY 818 CCTGTTTCCCTTCGGCGGGGAAGTCAT---AGTTAGGAGGACACCTCACACTTCTCTCAG 874
Db 457 AlaGluSerValAlaAsnAlaGluHisGlnArgLysGlnAspSerLysArgLeuSerThr 476
QY 875 GAGAAGACACACAACCATGACCTCAGCTGGGAC----- 910
Db 477 GluHisSerAsnValSerAspTyrProProAlaAspGlyTyrAlaTyrSerSerSerMet 496
QY 911 -----CCC 913
Db 497 TyrHisArgGlyProHisAlaAspHisGlyGluAlaAlaMetProPheLysSerThrPro 516
QY 914 AAC-----TCAGGCGCTCCAGCCCCAACCTGCCCCAGCCAGCCCTGAAATGCAA 964
Db 517 AsnHisGlnValAspArgAsnTyrAspAlaValLysThrGlnProLysLysTyrAlaLys 536
QY 965 GTCCTGTACAGCTTGAAGCTAGGAACCCACGGAGCTGACTGTGTGTCAGGAGAGAAG 1024
Db 537 SerLysTyrAspPheValAlaArgAsnSerSerGluLeuSerValMetLysAspVal 556
QY 1025 CTGGAGGTTCTGGACCACAGCAGCGGTGCTGGTGAAGAATGAGCGGACCGAGC 1084

```
Db 557 LeuGluIleLeuAspAspArgGlnTrpTrpLysValArgAsnAlaSerGlyAspSer 576
Qy 1085 GGCTACATTCACGACACACCTCGAG-
Db 577 GlyPheValProAsnAsnIleLeuAspIleMetArgThrProGluSerGlyValGlyArg 596
Qy 1111 -----
Db 597 AlaAspProProTyrThrHisThrIleGlnLysGlnArgThrGluTyrGlyLeuArgSer 616
Qy 1112 -----CCCTTA 1117
Db 617 AlaAspThrProSerAlaProSerProProProThrProAlaProValProValProLeu 636
Qy 1118 CAGCCGGGACCCCTGGG-----
Db 637 ProProSerValProAlaProValSerValProLysValProAlaAspValThrArgGln 656
Qy 1135 -----
Db 657 AsnSerSerSerAspSerGlySerIleValArgAspSerGlnArgTyrLysGln 676
Qy 1135 -----
Db 677 LeuProValAspArgArgLysSerGlnMetGluGluValGlnAspGluLeuPheGlnArg 696
Qy 1136 ---ACCAGGCCAGTCACCTCTCGG-----GTTCCA 1165
Db 697 LeuThrIleGlyArgSerAlaAlaGlnArgLysPheHisValProArgGlnAsnValPro 716
Qy 1166 ATGCTTCGACTTAGC-----TCGAGCCTGAAGAGTCCACAGACTGGCTGCAGGCAGAG 1219
Db 717 ValIleAsnIleThrTyrAspSerSerProGluGluValLysThrTrpLeuGlnSerLys 736
Qy 1220 AACTTCTCACTGCCAGGTGAGACACTTGGGTCCCTGACGGGAGCCAGCTACTTCGC 1279
Db 737 GlyPheAsnProValThrValAsnSerLeuGlyValLeuAsnGlyAlaGlnLeuPheSer 756
Qy 1280 ATAAGACCTGGGAGCTACAGACTATGTCACAGAGGAGCCACCACTCTCTCCCGG 1339
Db 757 LeuAsnLysAspGluLeuArgSerValLysPro---GluGlyAlaArgValPheAsnGln 775
Qy 1340 CTGGAGGCTGCAGAGGAGTGCTG 1363
Db 776 IleThrValGlnLysAlaAlaLeu 783
```

```
RESULT 4
US-08-931-820-4
; Sequence 4, Application US/08931820
; Patent No. 6010863
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Assay for collagen degradation
; NUMBER OF SEQUENCES: 4
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patencin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/931,820
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP 96202596.1
; FILING DATE:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1057 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; TISSUE TYPE: Collagen type III
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 1055
; OTHER INFORMATION: /label= Modified
; OTHER INFORMATION: /note= "Ala may be Pro"
US-08-931-820-4

Alignment Scores:
Pred. No.: 9,01e-09 Length: 1057
Score: 200,50 Matches: 157
Percent Similarity: 32,01% Conservative: 37
Best Local Similarity: 25,91% Mismatches: 224
Query Match: 6,41% Indels: 188
DB: 3 Gaps: 34

US-09-762-021A-1 (1-1710) x US-08-931-820-4 (1-1057)
Qy 10 ACTGAACACAGCCTGCAGAA---GGCTCTGGAGGAGAGAGCTGGAGCAAGACCTCGACT 66
Db 498 ThrGlyGlyProProGlyGluAsnGlyLysProGlyGluProGlyProLysGlyAspAla 517
Qy 67 TGGAGGCGCTTCAGCCAGCCAGGACAGATGGAGGGGCGCTGCTATGGAAGGCCGCTCCC 126
Db 518 GlyAlaProGlyAlaProGlyGlyLysAspAlaGlyAlaProGlyGluArgGlyPro 537
Qy 127 TATGGA-----GCAGGCACGCTATCTGAGCCGGGATCCCTCCAGAACAGACC 174
Db 538 ProGlyLeuAlaGlyAlaProGlyLeuArgGlyGlyAlaGlyProGly----- 554
Qy 175 CCACGAGACCCCTAGA-----GCACAGCTCCACACATCCCCAGGCCCTCCGACG 228
Db 555 ---ProGluGlyGlyLysGlyAlaAlaGlyProProGlyProGlyAlaAlaGlyThr 573
Qy 229 CCACACAGTGCCGCGAGAACCAAGTGCTTTACTCTGCTCCTCCCAAGGCGTCTCTTC 288
Db 574 Pro-----GlyLeuGlnGly 578
Qy 289 CCCGAGGACCCAGAGAGGAGGAGAGTGTGAACCATGTCTTAAGGACATTTGAGCT 348
Db 579 MetProGlyGluArgGlyGlyLeuGlySer-----ProGlyProLysGlyAsp----- 594
Qy 349 GTTCATGGGAAGCTGCAGAGAGGCCAGGCAAGAACACAGCAGGAGAGAAATTTGGGAA 408
Db 595 ---LysGlyGluProGly---GlyProGlyAlaAspGlyVal-----Pro 607
Qy 409 AAAAAACAAGGACCA---GGAGGTCTCACCCAGGCACAGTACATTGACTGCTTCCAGAA 465
Db 608 GlyLysAspGlyProArgGlyProThrGlyPro----- 618
Qy 466 GATCAAGTACAGCTTCAACCTCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAGTGC 525
Db 619 -----IleGlyProProGlyProAlaGlyGln-----ProGlyAspLysGly 632
Qy 526 CCCTGAGCTCGTACACATCTCTTCAAGTCCCTGAATTCATCTCTGCGCAGGTG---CCC 582
Db 633 GluGlyGly-----AlaProGlyLeu---ProGlyIleAlaGlyPro 645
Qy 583 TGAGGCTGGCTAGCAGCCCAAGTATCTCACCCCTCTCCACCCCTCAAGGATATCAACCT 642
Db 646 ArgGlySerProGlyGluArgGlyGluThrGlyProGlyProGlyProAlaGlyPheProGly 665
Qy 643 GCTACAGTCTGTCTAAGCCACCTGAGAGTAACTTTTGGATGGGTG----- 691
Db 666 AlaProGlyGlnAsn-GlyGluProGlyGlyLysGlyGluArgGlyAlaProGlyGluLys 685
Qy 692 -----GGCCAGCCTGGACCACTAGCGGGCCGACTGGACAGCGCATGAGCCCT 741
Db 685 sGlyGluGlyPro-ProGly---ValAlaGlyProProGlyGlySerGlyProAlaG 704
```

QY 742 GGCCTACCAACCCACATTTCTCAGATGACTGGCAACTTCAGAGCCCT----- 788
Db 704 lyProProGlyProGlnGly---ValLysGlyGluArgGlySerProGlyGlyProGlyA 723
QY 789 -----CCAGCCAAAGCAGCCCTTAGGATACCA 813
Db 723 laAlaGlyPheProGlyAlaAArgGlyLeuProGlyPro-ProGlySerAsnGlyAsnPro 742
QY 814 GGACCTGTTTCCCTTCGGCGGGGAAGTCATAGTTAGGGACCACTCACACTTTCCCTCA 873
Db 743 GlyProProGlyProSerGly-----SerPro 751
QY 874 GGAGAAGACACACAACTGACCTCA-----GCC 903
Db 752 GlyLysAspGlyProProGlyProAlaGlyAsnThrGlyAlaProGlySerProGlyVal 771
QY 904 TGGGAGACCCCACTCCAGGCCCTCCAGCCC-----CAAAACC 939
Db 772 SerGlyProLysGlyAspAlaGlyGlnProGlyGlyLysGlySerProGlyAlaGlnGly 791
QY 940 TGGCCAGCCAGCCCTCAAAATCCAAGTCTTGTACGAGTTTGAAGCTAGGACCCAGGGA 999
Db 792 ProProGlyAlaProGlyProLeuGlyIleAlaGlyIleThrGlyAlaArgGlyLeuAla 811
QY 1000 ACTGACTGTGTCCAGGGAGAGAGCTGGAGCTTCTGGACCAACAGCGGTGGTGGCT 1059
Db 812 GlyProProGlyMetProGlyProArgGlySerProGlyProGlnGly-----ValLys 829
QY 1060 GTTGAAGATGAGCC-----GGGACGGAGCGGCTACATTCCAAGCAACATCCTGGAGCC 1113
Db 830 GlyGluSer---GlyLysProGlyAlaAsnGlyLeuSerGlyGluArgGlyProProGly 848
QY 1114 CCT-----ACAGCCGGGGACCCC-----TGGGACCCA 1140
Db 849 ProGlnGlyLeuProGlyLeuAlaGlyThrAlaGlyGluProGlyArgAspGlyAsnPro 868
QY 1141 GGCCAGTCAACCTCTCGGCTTCCAAATGCTTCGACTTAGCTCGAGCGCTGAAGAGGTCAAC 1200
Db 869 GlySerAspGlyLeuProGlyArgAspGlySerProGlyGlyLysGlyAspArgGlyGlu 888
QY 1201 AGACTGGCTGCAGGCAGAGAACTTCTCCACTGCCAGGTGAGGACACTTGGTCCCTGAC 1260
Db 889 -----Asn 889
QY 1261 GGGAGCCAGCTACTTCGCATAGACCTGGGAGCTACAGATGCTATGTCACAGAGGC 1320
Db 890 GlySerPro-----GlyAlaProGlyAlaProGlyHisProGly 902
QY 1321 CCCACGAATCTGTCCGCTGGAGCTGTCCAGAGGATGCTGGGGAATAG----- 1371
Db 903 ProProGlyProValGlyProAlaGlyLysSerGlyAspArgGlyGluSerGlyProAla 922
QY 1372 ----CCCTTAGGCACCACTTACAGACCTCCAAAGACCAAGCCCGCTGTATGCAAGATGCG 1428
Db 923 GlyProAlaGlyAlaProGlyProAlaGlySerArgGlyAlaProGlyProGlnGlyPro 942
QY 1429 AGATCTGATACCCATTAGAGCCCGGAGAAATTCCTCTTCTGGATCCAG----- 1476
Db 943 ArgGlyAspLysGlyGluThrGlyGluArgGlyAlaAlaGlyIleLysGlyHisArgGly 962
QY 1477 TTTTGACGACCAACCC-----ACACCCAGCTCACACAGCAAAACAATGGACA 1524
Db 963 PheProGlyAsnProGlyAlaProGlySerPro----- 973
QY 1525 GGCCAGAGGCTGAACAAACAGTGT-----CCTTCTGCTGT 1563
Db 974 GlyProAlaGlyGlnGlnGlyAlaIleGlySerProGlyProAlaGlyProArgGlyPro 993
QY 1564 GTTGGAGCTCC 1575
Db 994 ValGlyProSer 997

RESULT 5
US-08-963-825-21
: Sequence 21, Application US/08963825
: Patent No. 6110689
: GENERAL INFORMATION:
: APPLICANT: Qvist, Per
: APPLICANT: Bonde, Martin
: TITLE OF INVENTION: A Method for Assaying Collagen Fragments
: TITLE OF INVENTION: In Body Fluids, A Test Kit and Means for Carrying Out the
: TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
: TITLE OF INVENTION: Disorders Associated with the Metabolism of
: NUMBER OF SEQUENCES: 21
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Darby & Darby PC
: STREET: 805 Third Avenue
: CITY: New York
: STATE: New York
: COUNTRY: USA
: ZIP: 10022
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/963,825
: FILING DATE:
: CLASSIFICATION: 436
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US/08/187,319
: FILING DATE: 21-JAN-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Gogoris, Adda C
: REGISTRATION NUMBER: 29,714
: REFERENCE/DOCKET NUMBER: 4305/08701
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 212-527-7700
: TELEFAX: 212-753-6237
: TELEX: 236687
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1078 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: ORIGINAL SOURCE:
: ORGANISM: Homo sapiens
: IMMEDIATE SOURCE:
: CLONE: COLLAGEN ALPHA 1 (III)
US-08-963-825-21

Alignment Scores:
Pred. No.: 5,78e-08 Length: 1078
Score: 191.00 Matches: 147
Percent Similarity: 29.11% Conservative: 37
Best Local Similarity: 23.26% Mismatches: 229
Query Match: 6.11% Indels: 219
DB: 3 Gaps: 26

US-09-762-021A-1 (1-1710) x US-08-963-825-21 (1-1078)
QY 7 GCGACTGAAGACCGCTCGACAGGCTCTGGAGGAGCTGGAGCAAGACCTCGACT 66
Db 334 AlaAlaGlyGluArgGlyAlaLeuGlySerArgGlyProAlaGly----- 348
QY 67 TGGAGCCTTACAGCGCCAGGACAGATGGAGGGGCTGTATGGAAGCCGCTCC 126
Db 349 -----ProAsnGlyIleProGlyGlu---LysGlyProAlaGlyGluArgGlyAlaPro 365
QY 127 TATGAGCAGCAGCCCTATCTGGAGCCGGGATCTCCAGAACAGCCGCCAGGAGAC 186
Db 366 GlyProAlaGlyProArgGlyAlaAlaGlyGluProGlyArgAspGlyValProGlyGly 385


```
OY 187 CTTAGAGCACAGCCTCCACATCCCAAGGCCCTGCCAGCCACACAGTGCCCGAGA 246
   |||
   :
Db 386 ProGlyMetArg-----GlyMetProGlySerProGlyGlyProGlySerAspGlyLys 403
   :
OY 247 ACCAAGTGCCTTTACTGTGCTCTCCCAAGCGGCTCTTCCCGGAGGAGCCAGAGAG 306
   |||
Db 404 ProGlyProProGlySerGlnGlySerGlyArgProGlyProGlyProSerGly 423
   |||
OY 307 GGACGAGGAAGT-----GCTGAACCATGCTCCTAAGGACATTTGAGCTGTTTCAT 354
   |||||
Db 424 ProArgGlyGlnProGlyValMetGlyPheProGlyProGlyProGlyAsnAspGlyAlaPro 443
OY 355 GGGAAAGCTGGAGAA-----GGCCAGGCAAGACAGCAGGAGGAAGAAATTTGGGAA 408
   |||||
Db 444 GlyLysAsnGlyGluArgGlyGlyProGly----- 453
OY 409 AAAAAAAGGACAGGAGGAGTCTCACCCAGGACAGTACATTCGCTTCCAGAGAT 468
   |||||
Db 454 -----GlyProGly----- 459
OY 469 CAAGTACAGCTTCAACCTCCTGGGAAGGTGGCCACCTGGCTGAAGAGAGACAAGTGCCCC 528
   |||||
Db 460 -----ProProGlyLysAsnGlyGluTyrgly----- 468
OY 529 TGGCTCGTACACATCCTCTTCAAGTCCCTGAACCTTCATCCTGGCCAGGTGCCGTGAGGC 588
   |||
Db 469 -----ProGlnGlyProProGly-----ProThrGly 477
OY 589 TGGCTAGCAGCCCAAGTATGATCTCACCCCTCCTCACCCCTAAAGCTATCAACCTGCTACA 648
   |||
Db 478 ProGlyGlyAspLysGlyAspThrGlyPro-ArgGlyProGlnGlyLeuGlnGlyLeuPr 497
OY 649 GTCCTGTCTAAGCCACCTCAGAGTAAC----- 676
   |||||
Db 497 oGlyThrGlyClyProProGlyGluAsnGluLysProGlyClyProGlyProLysGlyG1 517
OY 676 ----- 676
Db 517 uAlaGlyAlaProGlyAlaProGlyGlyLysGlyAspAlaGlyAlaProGlyGluArgG1 537
OY 677 -----CATTGGATGGGTGGGCGCCAGCTGGACAC 708
   |||
Db 537 yProProGlyLeuAlaGlyAlaProGlyLeuArgGlyGlyAlaGlyPro-ProGlyPro- 556
OY 709 TAGCGGGCGGACTGGACAGCGATAGCCCTGCCCTACCAACCCACATTCCTCAGATGA 768
   |||
Db 557 --GluGlyGlyLysGlyAlaAlaGlyProProGlyPro-----ProGlyAlaAlaGlyT 574
OY 769 CTGCAAACTTCCAGAGCCCTCCAGCCCAAGCACCCCTTAGGATACCAGGACCCCTGTTCCCT 828
   |||
Db 574 hrProGlyLeuGlnGlyMetPro----- 581
OY 829 TCGCGGGGAAGTATAGTTAGGAGCAGCCTCACATTCCTCCTCAGGAGAGACACAA 888
   |||||
Db 582 -----GlyGluArgGlyGlyLeuGlySerProGlyPro----- 592
OY 889 CCATGACCCCTCAGCTGGGAGCCCA-----ACTCAGGCGCCCTC 927
   |||||
Db 593 -----LysGlyAspLysGlyClyProGlyProGlyGlyAlaAspGlyValProGlyLysA 611
OY 928 CAGCCCCAAACCTCCCGAGCCGCTGAGGAGGAGAGAGAGCTGGAGGTTCTGGACACACAA 1047
   |||||
Db 621 ly-ProProGlyProAlaGlyGlnProGlyAspLysGlyGlyGlyAlaProGlyLeu 640
OY 1048 GCGGTGTGCTGTGAAGATGAGCGGAGGAGCGGCTACATTCACAGACATCCT 1107
   |||||
Db 641 ProGlylleAlaGlyProArgGlySerProGlyGluArg----- 653
OY 1108 GGAGCCCTTACAGCGGGGAGCCCTGGGAGCCAGGCGCAGTCAACCTCTCGGGTTCCAAT 1167
```

```
Db 654 -----GlyGluThrGlyProProGlyProAlaGlyPheProGlyAlaPro 668
   |||
OY 1168 GCTTCGACTTAGCTCGAGGCTGAAGAGGTCACAGACTGGCTGCAGCAGAGAACTTCTC 1227
   :
Db 669 GlyGlnAsnGlyGluProGlyGlyLysGlyGluArgGlyAlaProGlyGlyLys----- 686
   :
OY 1228 CACTGCCACGCTGAGGACACTTGGTCCCTGACGGGAGGAGCAGTACTTGCATAGAACC 1287
   |||||
Db 587 -----GlyGlu-----GlyGlyProProGlyValAlaValProPro 698
OY 1288 TGGGAGCTACAGATGCT-----ATGTCC 1311
   |||
Db 699 GlyGlySerGlyProAlaGlyProProGlyProGlnGlyValLysGlyGluArgGlySer 718
OY 1312 ACAGGAGGCC----- 1323
Db 719 ProGlyGlyProGlyAlaAlaGlyPheProGlyAlaArgGlyLeuProGlyProGly 738
OY 1324 -----ACGAATCCTGTCGCGGCTGGAGGCTGTGAGAAGGATGCTGGGATAAGCCCTTA 1377
   |||||
Db 739 SerAsnGlyAsnProGlyProProGlyProSerGlySerProGlyLysAspGlyProPro 758
OY 1378 GGCACACGCTTAGACCTCCAGAACACAGGCCCGCTGATGCAAGATGGCAGATCTGAT 1437
   |||
Db 759 GlyProAlaGlyAsnThrGlyAlaProGlySerProGlyValSerGlyProLysGlyAsp 778
OY 1438 ACCATTAGAGCCCGAGAATTCCTCTTCTGGATCCAGTTTGCAGCAAAAC---CCCA 1494
   |||
Db 779 AlaGlyGlnProGlyGlyLysGlySerProGlyAlaGlnGlyProProGlyAlaProGly 798
OY 1495 CCC-----CAGCTCACACAGCAAAACAATGGACAGCCCGCAGAGCT----- 1536
   |||
Db 799 ProLeuGlylleAlaGlylleThrGlyAlaArgGlyLeuAlaGlyProProGlyMetPro 818
OY 1537 -----GAAGCAAAACAGTGTCTTCTGGCTGTGTTGGAGCCTCCCGACAGTACACCAT 1590
   |||
Db 819 GlyProArgGlySerProGlyProGlnGlyValLysGlyGluSer----- 833
OY 1591 TTATTTTACCTCTTCCCAACCTGGAGCA 1620
   |||||
Db 834 -----GlyLysProGlyAla 838
```

RESULT 6

```
US-09-500-811-21
; Sequence 21, Application US/09500811
; Patent No. 632314
; GENERAL INFORMATION:
; APPLICANT: Qvist, Per
; APPLICANT: Bonde, Martin
; TITLE OF INVENTION: A Method for Assaying Collagen Fragments
; TITLE OF INVENTION: in Body Fluids, A Test Kit and Means for Carrying Out the
; TITLE OF INVENTION: Method and Use of the Method to Diagnose the Presence of
; TITLE OF INVENTION: Disorders Associated with the Metabolism of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby & Darby PC
; STREET: 805 Third Avenue
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/500,811
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/187,319
```

; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Gogolis, Adda C
 ; REGISTRATION NUMBER: 29,714
 ; REFERENCE/DOCKET NUMBER: 4305/08701
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-527-7700
 ; TELEFAX: 212-753-6237
 ; TELEX: 236687
 ; INFORMATION FOR SEQ ID NO: 21:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 1078 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; ORIGINAL SOURCE:
 ; ORGANISM: Homo sapiens
 ; IMMEDIATE SOURCE:
 ; CLONE: COLLAGEN ALPHA 1 (III)
 ; US-09-500-811-21

Alignment Scores:
 Pred. No.: 5,78e-08 Length: 1078
 Score: 191.00 Matches: 147
 Percent Similarity: 29.11% Conservative: 37
 Best Local Similarity: 23.26% Mismatches: 229
 Query Match: 6.11% Indels: 219
 DB: 4 Gaps: 26

US-09-762-021a-1 (1-1710) x US-09-500-811-21 (1-1078)

```

; 7 GCGACTGAAGACACCGCTGCAAGAGCTGTGGAGGAGCTGGAGCAAGACCTCGACT 66
;      |||      |||      |||      |||      |||      |||      |||
; 334 AlaAlaGlyGluArgGlyAlaLeuGlySerArgGlyProAlaGly----- 348
;
; 67 TGGAGCCTTCAGCAGCGGCGAGCAGATGAGAGGGGCTGTATGGAAGCCGCTCC 126
;      |||      |||      |||      |||      |||      |||      |||
; 349 -----ProAsnGlyIleProGlyGlu---LysGlyProAlaGlyGluArgGlyAlaPro 365
;
; 127 TATGAGCAGGACGCTACTGTGAGCGGGGATCTCCAGAACAGCCCGCAGGAGAC 186
;      |||      |||      |||      |||      |||      |||      |||
; 366 GlyProAlaGlyProArgGlyAlaAlaGlyGluProGlyArgAspGlyValProGlyGly 385
;
; 187 CCTAGACACACCTCCACCATCCCAAGGCCCTGCCACGCCACACAGTGCCTCCGAGA 246
;      |||      |||      |||      |||      |||      |||      |||
; 386 ProGlyMetArg-----GlyMetProGlySerProGlyGlyProGlySerAspGlyLys 403
;
; 247 ACCAAGTCCTTACTCTGCTCCTCCAAAGGGGCTCTCTTCCCGGAGGACCCAGAGAG 306
;      |||      |||      |||      |||      |||      |||      |||
; 404 ProGlyProProGlySerGlnGlyGluSerGlyArgProGlyProGlyProSerGly 423
;
; 307 GCACGAGGAAGT-----GCTGAACCATCTCTTAAGGACATTGACCTGTCAT 354
;      |||      |||      |||      |||      |||      |||      |||
; 424 ProArgGlyGlnProGlyValMetGlyPheProGlyProGlyProGlyValAspGlyAlaPro 443
;
; 355 GGGAAAGCTGGAGAA-----GGCCCCAGCAAGACAGCAGGAAAGAAATTTGGGAA 408
;      |||      |||      |||      |||      |||      |||      |||
; 444 GlyLysAsnGlyGluArgGlyGlyProGly----- 453
;
; 409 AAAAAACAGGACGAGGAGGTCTCACCCAGGCACAGTACATTGACTGCTCCAGAAAGAT 468
;      |||      |||      |||      |||      |||      |||      |||
; 454 -----GlyProGly-----ProGlnGly 459
;
; 469 CAAGTACAGCTTCAACCTCTGGGAGGCTGGCCACCTGGCTGAAGAGACAAAGTGCCCC 528
;      |||      |||      |||      |||      |||      |||      |||
; 460 -----ProProGlyLysAsnGlyGluTyrGly----- 468
;
; 529 TGAGCTCGTACATCTCTTCAAGTCCCTGAACCTTCATCTGGCCAGGTGCGCTGAGGC 588
;      |||      |||      |||      |||      |||      |||      |||
; 469 -----ProGlnGlyProProGly-----ProThrGly 477
;
; 589 TGGCCTAGCAGCCCAAGTGATCTACCCCTCTCACCCCTAAAGTATCAACCTCTGACA 648
;      |||      |||      |||      |||      |||      |||      |||

```

```

Db 478 ProGlyGlyAspLysGlyAspThrGlyPro-ArgGlyProGlnGlyLeuGlnGlyLeuPr 497
;
; 649 GTCCCTGTCTAAGCCCAACCTGTAGAGTAAC----- 676
;      |||      |||      |||      |||      |||      |||      |||
; 497 OGlyThrGlyGlyProProGlyGluAsnGlnLysProGlyGluProGlyProLysGlyGly 517
;
; 676 ----- 676
;
; 517 uAlaGlyAlaProGlyAlaProGlyGlyLysGlyAspAlaGlyAlaProGlyGluArgGly 537
;
; 677 -----CTTTGGATGGGTTGGCCCGCAGCCCTGGACCAC 708
;      |||      |||      |||      |||      |||      |||      |||
; 537 yProProGlyLeuAlaGlyAlaProGlyLeuArgGlyGlyAlaGlyPro- 556
;
; 709 TAGCGGGCGGCGACTGACAGAGGCGATGAGCCCTGCCCTACCAACCCACATCTCAGATGA 768
;      |||      |||      |||      |||      |||      |||      |||
; 557 --GluGlyGlyLysGlyAlaAlaGlyProProGlyPro-----ProGlyAlaAlaGlyT 574
;
; 769 CTGGCAACTTCCAGAGCCCTCCAGCAAGCACCCCTTAGGATACCAGGACCCTGTTTCCCT 828
;      |||      |||      |||      |||      |||      |||      |||
; 574 hrProGlyLeuGlnGlyMetPro----- 581
;
; 829 TCGGGCGGGAAGTCAATAGTTTAGGAGCAGCCTCACACTTCTCAGGAGACACACAA 888
;      |||      |||      |||      |||      |||      |||      |||
; 582 -----GlyGluArgGlyLeuGlySerProGlyPro----- 592
;
; 889 CCATGACCTCAGCCTGGGAGCCCA-----ACTCCAGGCCCTC 927
;      |||      |||      |||      |||      |||      |||      |||
; 593 -----LysGlyAspLysGlyGluProGlyGlyProGlyAlaAspGlyValProGlyLysA 611
;
; 928 CAGCCCAAACTGCCAGCAGCCCTGAAAATGCAAGTCTTGTACGAGTTTGAAGCTAG 987
;      |||      |||      |||      |||      |||      |||      |||
; 611 spGlyProArgGlyProThrGlyPro-----IleG 621
;
; 988 GAACCCAGGGAAGTACTGTGGTCCAGGAGAGAAAGCTGGAGGTTCTGGACCACAGCA 1047
;      |||      |||      |||      |||      |||      |||      |||
; 621 ly-ProProGlyProAlaGlyGlnProGlyAspLysGlyGlyAlaProGlyLeu 640
;
; 1048 GCGGTGGTGGCTGTGAAGAATGAGCGGGAGCGGCTACATTCCAAGCAACATCCT 1107
;      |||      |||      |||      |||      |||      |||      |||
; 641 ProGlyIleAlaGlyProArgGlySerProGlyGluArg----- 653
;
; 1108 GGAGCCCTACAGCCGGGACCCCTGGAGCCAGCCAGTCACCTCTCGGTTTCCAAT 1167
;      |||      |||      |||      |||      |||      |||      |||
; 654 -----GlyGluThrGlyProGlyProAlaGlyPheProGlyAlaPro 668
;
; 1168 GCTTCGACTTACCTCGAGCCCTGAAGAGTCAACACTGGCTGCGAGGAGAGAACTTCTC 1227
;      |||      |||      |||      |||      |||      |||      |||
; 659 GlyGlnAsnGlyGluProGlyGlyGlyGluArgGlyAlaProGlyGlyLys----- 686
;
; 1228 CACTGCCACGGTGAGGACACTTGGGTCCCTGACGGGAGCCAGCTACTTCGCATAAGACC 1287
;      |||      |||      |||      |||      |||      |||      |||
; 687 -----GlyGlu-----GlyGlyProProGlyValAlaValProPro 698
;
; 1288 TGGGAGGTACAGATGCT-----ATGTCC 1311
;      |||      |||      |||      |||      |||      |||      |||
; 699 GlyGlySerGlyProAlaGlyProProGlyProGlnGlyValLysGlyGluArgGlySer 718
;
; 1312 ACAGAGGCCCC----- 1323
;      |||      |||      |||      |||      |||      |||      |||
; 719 ProGlyGlyProGlyAlaAlaGlyPheProGlyAlaArgGlyLeuProGlyProGly 738
;
; 1324 -----ACGAATCTTGTCCCGCTGAGGCTGTCAAGAGGATGCTGGGATAAAGCCCTTA 1377
;      |||      |||      |||      |||      |||      |||      |||
; 739 SerAsnGlyAsnProGlyProProGlyProSerGlySerProGlyLysAspGlyProPro 758
;
; 1378 GGCACCACTTAGACACCTCCAAAGAACCCAGGCCGCTGTATGCAAGATGGCAGATCTGAT 1437
;      |||      |||      |||      |||      |||      |||      |||
; 759 GlyProAlaGlyAsnThrGlyAlaProGlySerProGlyValSerGlyProLysGlyAsp 778
;
; 1438 ACCCATAGACCCCGAGAAATTCCTTCTGATCCAGTTTGCACGAAAC---CCACA 1494
;      |||      |||      |||      |||      |||      |||      |||
; 779 AlaGlyGlnProGlyGlySerProGlyAlaGlnGlyProProGlyAlaProGly 798

```



```
Db 454 -----GlyProGly-----ProGlnGly 459
Qy 469 CAAGTACAGCTTCAACCTCTCGGAAGGCTGGCCACCTGGCTGAAGAGACAAAGTGGCCCC 528
Db 460 -----ProProGlyLysAsnGlyGluThrGly----- 468
Qy 529 TGAGCTCGTACACATCTCTTCAAGTCCCTGAACCTTCATCTCGCCAGGTCGCCAGGC 588
Db 469 -----ProGlnGlyProProGly-----ProThrGly 477
Qy 589 TGGCTAGCAGCCCAAGTATCATCTCAGCTTCAACCTTAACCTATCAACCTGCTACA 648
Db 478 ProGlyGlyAspLysGlyAspThrGlyPro-ArgGlyProGlnGlyLeuGlnGlyLeuPr 497
Qy 649 GTCCTGCTTAAGCCCACTCGAGATAAC----- 676
Db 497 oGlyThrGlyGlyProProGlyGluAsnGluLysProGlyGluProGlyProLysGlyG 517
Qy 676 ----- 676
Db 517 uAlaGlyAlaProGlyAlaProGlyLysGlyAspAlaGlyAlaProGlyGluArgG 537
Qy 677 -----CTTTGGATGGGTGGGCCCGCCAGCTGGACAC 708
Db 537 yProProGlyLeuAlaGlyAlaProGlyLeuArgGlyGlyAlaGlyPro-ProGlyPro- 556
Qy 709 TAGCCGGCGGACTGGACAGCATGAGCCCTGCTCCCTACCAACCCACATCTCAGATGA 768
Db 557 --GluGlyGlyLysGlyAlaGlyProProGlyPro-----ProGlyAlaGlyT 574
Qy 769 CTGGCAACTCCAGAGCCCTCCAGCCCAAGCACCCCTTAGGATACAGGACCCCTGTTCCCT 828
Db 574 hrProGlyLeuGlnGlyMetPro----- 581
Qy 829 TCGCGGGGAGTATAGTTAGGAGCACTCACATCTTCTCAGGAGAGACACAA 888
Db 582 -----GlyGluArgGlyGlyLeuGlySerProGlyPro----- 592
Qy 889 CCATGACCTCAGCTGGGACCCCA-----ACTCCAGGCCCTC 927
Db 593 -----LysGlyAspLysGlyGluProGlyProGlyAlaAspGlyAlaProGlyLys 611
Qy 928 CAGCCCAAACTCCAGCAGCAGCCCTGAAATGCAAGTCTTGACGAGTTGAAGCTAG 987
Db 611 spGlyProArgGlyProThrGlyPro-----IleG 621
Qy 988 GAACCCACGGGAAGTACTGTGTCCAGGAGAGAGCTGGAGTTCTGGACCAACAGCAA 1047
Db 621 ly-ProProGlyProAlaGlyGlnProGlyAspLysGlyGlyGlyAlaProGlyLeu 640
Qy 1048 GCGGTGTGTGTGAAGATAGGCGGAGCGGCTACATTCACAGCAACATCCT 1107
Db 641 ProGlyLeaAlaGlyProArgLysProGlyGluArg----- 653
Qy 1108 GGAGCCCTACAGCGGGGACCCCTGGGACCCAGGCGGCTGCTCGGTTCCAAT 1167
Db 654 -----GlyGluThrGlyProProGlyProAlaGlyPheProGlyAlaPro 668
Qy 1168 GCTTCGACTTAGCTCGAGGCTGAAGAGGTACAGACTGGCTGCAGCAGAGAACTTCTC 1227
Db 669 GlyGlnAsnGlyGluProGlyGlyLysGlyGlyArgGlyAlaProGlyGluLys----- 686
Qy 1228 CACTGCCACGGTAGGACACTTGGTCCCTGACGGGAGCCAGCTACTTCGCATAAGACC 1287
Db 687 -----GlyGlu-----GlyGlyProGlyValAlaValProPro 698
Qy 1288 TGGGAGCTACAGATGCT-----ATGTCC 1311
Db 699 GlyGlySerGlyProAlaGlyProProGlyProGlnGlyValLysGlyGluArgGlySer 718
Qy 1312 ACAGGAGGCC----- 1323
```

```
Db 719 ProGlyGlyProGlyAlaAlaGlyPheProGlyAlaArgGlyLeuProGlyProProGly 738
Qy 1324 -----ACGAATCTCTCCGCGCTGGAGGCTGTGTCAGAAGGATGCTGGGGATAAGCCCTTA 1377
Db 739 SerAsnGlyAsnProGlyProProGlyProSerGlySerProGlyLysAspGlyProPro 758
Qy 1378 GGCACACAGCTTAGACACTCCAAGAACACAGCCCGCTGATGCAAGATGCGCAGATCTGAT 1437
Db 759 GlyProAlaGlyAsnThrGlyAlaProGlySerProGlyValSerGlyProLysGlyAsp 778
Qy 1438 ACCATTAGAGCCCGAGAAATTCCTCTTCTGGATCCAGTTTCAGCAAAAC---CCACACA 1494
Db 779 AlaGlyGlnProGlyGluLysGlySerProGlyAlaGlnGlyProProGlyAlaProGly 798
Qy 1495 CCC-----CAGCTCACACAGCAAAACAATGGACAGCCCGAGGCT----- 1536
Db 799 ProLeuGlyLeuAlaGlyIleThrGlyAlaArgGlyLeuAlaGlyProProGlyMetPro 818
Qy 1537 -----GAAGCAACAGTGTCCCTTCTGGTGTGTTGGAGCTCCCGCAAGTAAACACCTAT 1590
Db 819 GlyProArgGlySerProGlyProGlnGlyValLysGlyGluSer----- 833
Qy 1591 TTATTTTACCTCTTCCCAACCTGGAGCA 1620
Db 834 -----GlyLysProGlyAla 838

RESULT 9
US-08-642-255-52
; Sequence 52, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BTR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 561 amino acids
; . TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-642-255-52

Alignment Scores: 2.85e-07 Length: 561
Pred. No.: 181.50 Matches: 131
Score: 31.03% Conservative: 13
Percent Similarity:
```

Best Local Similarity: 28.23% Mismatches: 204
Query Match: 5.80% Indels: 116
DB: 1 Gaps: 27

US-09-762-021a-1 (1-1710) x US-08-642-255-52 (1-561)

QY 25 GCAGAAAGGCTCTGGAGAGAGCTGGAGCAAGACCTCGACTTGGAGGCTTCAGCC--- 81
DB 188 AlaProGlyProAlaGlyProProGlyAlaProGlyProAlaGlyProProGlyAlaPro 207
QY 82 AGCCAGGACAGATGAGGGGCGCTCTATGGAAGGGCGCTCCCTATGGAGCA---GGC 138
DB 208 GlyProAlaGlyProProGlyAla---ProGlyProAlaGlyProProGlyAlaProGly 226
QY 139 ACGCTATCTGGAGCGGGGATCCCTCCAGAACAGCCACAGGAGGAGCCCT---AGAGCA 195
DB 227 ProAlaGlyProProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAla 246
QY 196 CAGCCTCCACCATCCCAAGGCCCTGCCAGCCACACAGTGGCCGAGAACCAAGTGC 255
DB 247 GlyProProGlyAlaPro---GlyProAlaGlyProProGlyAlaProGly----- 262
QY 256 CTTTACTCTGCTCTCCCAAGCGGTCTCTTCCCGGAGGAGCCAGAGAGGAGGAGCA 315
DB 263 -----ProAlaGlyProProGlyAlaHisGlyProAlaGlyProLysGlyAlaHisGly 280
QY 316 AGTGTCTGAACCA-----TGCTTAAGGACATGAGCTGTTTCATGGGAAAGCT 363
DB 281 ProAlaGlyProLysGlyAlaHisGlyProAlaGlyProLysGlyAlaHisGlyProAla 300
QY 364 GGAGAGGCCCGGCAAGACAGCAGGAGGAAGAANAATTTGGGAAAAAACAAGGACCA 423
DB 301 Gly-----ProLysGlyAlaPro 306
QY 424 GGGAGTCTACCCAGGCACAGTACATGACTGCTTCCAGAGATCAAGTACAGCTTCAA 483
DB 307 GlyProAlaGlyPro-----ProGlyAlaProGlyProAlaGly 319
QY 484 CTCTCTGGGAAGCTGGCCACCTG-----GCTGAAGGAGACAAAGTGGCCCTGA 531
DB 320 ProProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAlaGlyProPro 339
QY 532 GCTGCTACACATCTCTTCAAGTCCCTGAACCTCATCTGCGCCAGGTGCCCTGAGGCTGG 591
DB 340 GlyAlaProGlyProAlaGlyProProGlyAla---ProGly-----ProAlaGlyPro 356
QY 592 CTTAGCAGCCCAAGTATCTCACCCCTCTCACCCCTTAAGATATCAACCTGCTACAGTC 651
DB 357 ProGlyAlaProGlyProAlaGlyProProGly-----AlaProGlyProAlaGlyPro 374
QY 652 CTGCTAAGCCCACTGACAGTAACCTTTGGATGGGTTGGCCAGCCTGGAGCA--- 707
DB 375 ProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAla---GlyProProG 394
QY 708 --CTAGCCGGCGAGTGGACAGGCGATGAGCCCTGCCCTACCAACCCCAATCTTCAGA 765
DB 394 lyAlaProGlyProAlaGlyProProGlyAlaProGlyProAlaGlyProProGly---- 412
QY 766 TGACTGGCAACTCCAGAGCCCT---CCAGCCAAGCACCCCTTAGGATACCCAGGACCCCTGT 822
DB 413 -----AlaProGlyProAlaGlyProPro-GlyAlaProGlyProAla 426
QY 823 TTCCCTTCGGCGGGGAGTCATAGTTAGGGAGCACCTCACACTTTCCTCAGGAGAAC 882
DB 427 GlyProProGlyAlaProGly-----ProAlaGlyProPro 438
QY 883 ACACAACTATGACCTCAGCCCTGGGAGCCCACTCCAGGCCCTCCAGCCCAACACCTGC 942
DB 439 GlyAlaProGlyProAla-----GlyProPro---GlyAlaProGlyProAlaGlyPro 455
QY 943 CCAGCCAGCCCTGAANATCAAGTCTTTACAGAGTTTGAAGCTAGAACCCACGGGAAGCT 1002
DB 456 ProGlyAlaProGlyProAlaGly-----ProProGlyAla 467

QY 1003 GACTGTGCTCAGGAGAGAAAGCTGGAGGTTCTTGACACACAGCAAGCGTGTGCTGCTGT 1062
DB 468 Pro---GlyProAlaGlyProProGlyAlaProGlyPro-----AlaGly 481
QY 1063 GAAGAATGAGCGGGAGCGGAGCGGTACATCCCAAGCAACATCTCTGGAGCCCT---ACA 1119
DB 482 Pro-----ProGlyAlaProGlyPro 488
QY 1120 GCCGGGACCCTGGGACCCAGGCGCCAGTACCCTCTCGGGTTCCAATGCTTCGACTTAG 1179
DB 489 AlaGlyProProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAla--- 507
QY 1180 CTCGAGGCTGAAGAGGTACAGAGTGGCTGCAGGAGCAAGAACTTCTCCACTGCCACGCT 1239
DB 508 -----GlyProGlyAlaProGly----- 514
QY 1240 GAGGACACTTGGTCCCTGACGGGAGCCAGTACTTCGCATAAGACCTGGGAGCTACA 1299
DB 515 -----ProAlaGlyProProGly---AlaHisGlyProAlaGlyProLys 528
QY 1300 GATGCTATGTCCACAGAGGAGGAGGAGGAGGAGTCTGTCGCCGCTGGAGGCTGTCAAGAGAT 1359
DB 529 GlyAlaHisGlyProAlaGlyProLysGlyAlaMetAspProGlyArgTyroGlnLeuSer 548
QY 1360 GCTGGG 1365
DB 549 AlaGly 550
RESULT 10
US-09-219-849-3
; Sequence 3, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWMSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WESTEN, MARC W.T.
; APPLICANT: WIND, RICHELIE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE OF INVENTION: PREPARATION THEREOF
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3
; LENGTH: 504
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
; OTHER INFORMATION: amino acid sequence
US-09-219-849-3

Alignment Scores:
Pred. No.: 4,04e-07 Length: 504
Score: 179.50 Matches: 130
Percent Similarity: 29.64% Conservative: 9
Best Local Similarity: 27.72% Mismatches: 184
Query Match: 5.74% Indels: 146
DB: 4 Gaps: 27

US-09-762-021a-1 (1-1710) x US-09-219-849-3 (1-504)

QY 25 GCAGAAAGGCTCTGGAGGAGAGAGCTGGAGCAAGACCTCGACTTGGAGGCTTCAGCC--- 81
DB 155 AlaProGlyProAlaGlyProProGlyAlaProGlyProAlaGlyProProGlyAlaPro 174

```
Qy 82 AGGCCAGACAGATGGAGGGCCCTGCTATGTGAAGGCCGCTCCTATGTGAGCA---GGC 138
Db 175 GlyProAlaGlyProProGlyAla---ProGlyProAlaGlyProProGlyAlaProGly 193
Qy 139 ACGTATCTGAGCGCGGGATCCCTCCAGACACAGCCACAGAGGACCT---AGAGCA 195
Db 194 ProAlaGlyProProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAla 213
Qy 196 CAGCCTCCACCATCCCAAGGCCCTGCCACGCCACACCAAGTCCCGAGAACCAAGTGC 255
Db 214 GlyProProGlyAlaPro---GlyProAlaGlyProProGlyAlaProGly----- 229
Qy 256 CTTTACTCTGCTCCTCCAAAGCGGTCTCTTCCCGAGAGCCAGAGGAGGACGAGGA 315
Db 230 -----ProAlaGlyProProGlyAlaHisGlyProAlaGlyProLysGlyAlaHisGly 247
Qy 316 AGTCTGNACCA-----TGTCCTAAGGAGCATTCAGCTGTTTCATGGGAAAGCT 363
Db 248 ProAlaGlyProLysGlyAlaHisGlyProAlaGlyProLysGlyAlaHisGlyProAla 267
Qy 364 GGAGAAGGCCAGGCAAGACAGCAGGAGGAAGAAATTTGGGAAAAAACAAGGACCA 423
Db 268 Gly-----ProLysGlyAlaPro 273
Qy 424 GGGAGGTCTCACCAGGCACAGTACATTGCTGCTCCAGAGATCAAGTACAGCTTCAA 483
Db 274 GlyProAlaGlyPro-----ProGlyAlaProGlyProAlaGly 286
Qy 484 CCTCTGGGAGGCTGGCCACCTG-----GCTGAAGGAGACAAAGTGCCCTCGTA 531
Db 287 ProProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAlaGlyProPro 306
Qy 532 GCTCGTACACATCTCTTCAAGTCCCTGAAGTTCATCTGCTGAGGTCCTGAGGCTGG 591
Db 307 GlyAlaProGlyProAlaGlyProProGlyAla---ProGly-----ProAlaGlyPro 323
Qy 592 CCTAGCAGCCCAAGTATCTCACCCCTCCTCACCCCTAAAGCTATCAACCTGTACAGTC 651
Db 324 ProGlyAlaProGlyProAlaGlyProProGly-----AlaProGlyProAlaGlyPro 341
Qy 652 CTGTCTAAGCCCACTGAGAGTACCTTTGGATGGGTGGGCGCCAGCCGTCGACCA---- 707
Db 342 ProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAla---GlyProProG 361
Qy 708 --CTAGCGCGCCGACTGACGAGCGATGAGCCCTGCTCCCTACCAACCCACATTCCTAGA 765
Db 361 lyAlaProGlyProAlaGlyProProGlyAlaProGlyProAlaGlyProProGly---- 379
Qy 766 TGACTGGCAACTTCCAGAGCCCT---CCAGCCAAGCACCCCTTAGGATACCAAGGACCTGT 822
Db 380 -----AlaProGlyProAlaGlyProPro-GlyAlaProGlyProAla 393
Qy 823 TTCCTTTGGCGGGAAGTCATAGTTAGGAGGACACCTCACACTTCTCAGGAGAGAC 882
Db 394 GlyPro----- 395
Qy 883 ACACAACCATGACCTCAGCCTGGGAGCCCAACTCCAGGCGCTCCAGCCCAACCTGC 942
Db 396 -----ProGlyAlaProGlyProAlaGlyProProGlyAlaProGly----- 409
Qy 943 CCAGCCAGCCCTGAAATGCAAGTCTTTGTACGAGTTTGAAGCTAGGAACCCACCGGAAC 1002
Db 410 ProAlaGlyProProGlyAlaPro----- 417
Qy 1003 GACTGTGTCAGGAGAGAGCTGGAGGTCTGGACACAGCAAGCGGTGGTGGTGGT 1062
Db 418 -----GlyProAlaGlyProProGlyAlaProGlyPro-----AlaGly 430
Qy 1063 GAAGAAATGAGCGGAGCGGCTACATTCCAAGCAACATCCTGGAGCCCT---ACA 1119
Db 431 Pro-----ProGlyAlaProGlyPro 437
Qy 1120 GCCGGGACCCCTGGGAGCCAGGCGCAGTCAACCTCTCGGTTCCAAATGCTTCGACTTAG 1179
```

```
Db 438 AlaGlyProProGlyAlaProGlyProAlaGlyProProGlyAlaProGly----- 454
Qy 1180 CTCGAGCGCTCAAGAGGTCACAGACTGGCTGCAGGCAGAGAACTTCTCCACTGCCACGGT 1239
Db 455 -----ProAlaGly----- 457
Qy 1240 GAGGACACTTGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGACCTGGGGAGCTACA 1299
Db 458 -----ProGlyAlaProGlyProAla-----GlyProPro 468
Qy 1300 GATGCTATGTCCACAGGAGGCCCCACGAAT---CCTGTCCCGGCTGGA-----GGCTGT 1350
Db 469 GlyAlaProGlyProAlaGlyProProGlyAlaProGlyProAlaGlyProProGlyAla 488
Qy 1351 CAGAAGATGCTGGGATAAG 1371
Db 489 HisGlyProAlaGlyProLys 495
RESULT 11
US-09-219-849-7
; Sequence 7, Application US/09219849
; Patent No. 6150081
; GENERAL INFORMATION:
; APPLICANT: VAN HEERDE, GEORGE V.
; APPLICANT: VAN RIJN, ALEXIS C.
; APPLICANT: BOWMSTRA, JAN B.
; APPLICANT: DE WOLF, FREDERIK A.
; APPLICANT: MOOBROEK, ANDREAS
; APPLICANT: WERTEN, MARC W.T.
; APPLICANT: WIND, RICHELE D.
; APPLICANT: VAN DEN BOSCH, TANJA J.
; TITLE OF INVENTION: SILVER HALIDE EMULSIONS WITH RECOMBINANT COLLAGEN
; TITLE OF INVENTION: SUITABLE FOR PHOTOGRAPHIC APPLICATION AND ALSO THE
; FILE REFERENCE: 2728-2
; CURRENT APPLICATION NUMBER: US/09/219,849
; CURRENT FILING DATE: 1998-12-23
; NUMBER OF SEQ ID NOS: 50
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 7
; LENGTH: 552
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Illustrative
; OTHER INFORMATION: amino acid sequence
US-09-219-849-7
Alignment Scores:
Pred. No.: 5,61e-07 Length: 552
Score: 178.00 Matches: 153
Percent Similarity: 29.93% Conservatives: 26
Best Local Similarity: 25.59% Mismatches: 264
Query Match: 5.69% Indels: 155
DB: 4 Gaps: 33
US-09-762-021a-1 (1-1710) x US-09-219-849-7 (1-552)
Qy 22 CTGCAAGAGGCTCTGGAGGAAGAGCTGGAGCAAGACCTCGACTTGGAGGCGCTTCAGCC 81
Db 57 ProAlaGlyProGlyGlySerArgAspProGlyProGlyAlaGlnGlyProAlaGly 76
Qy 82 AGGCCAGACAGATGGAGGGCCCTGCTATGTGAAGGCCGCTCCTATGTGAGGAGCAGC 141
Db 77 ProGlyGlySerArgAspProGlyProGlyAlaGlnGlyProAlaGlyProGlyGly 96
Qy 142 CTATCTGGAGCGGGGATCCCTCCAGAACAGCCCA----- 177
Db 97 SerArgAspProGlyProProGlyAlaGlnGlyProAlaGlyProGlyGlySerArgAsp 116
Qy 178 CCAGGAGCCCTTAGAGCAGCAGCTCCCAACCATCCCAAG-----GCCCTGCCAGC 228
```

Qy	1243	GACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCCGATATAGACACCTGGGGAGCTACAGAT	1302
Db	419	-----ProAlaGly---ProLysGlyYAlaHisGlyProAlaGlyProLysGly	433
Qy	1303	GCTATGTCACAGGAGGCC-----ACGAATCTCTGTCCCGGCTGGA-----	1344
Db	434	AlaGlnGlyProAlaGlyProGlyGlySerArgAspProGlyProGlyYalaGlnGly	453
Qy	1345	-----GGCTGTGCAGAGGATGCTGGGGATAGCCCTTATGGCACCAGCTTATAGC	1392
Db	454	ProAlaGlyProGlyGlySerArgAspProGly-----ProProGlyYala-----	468
Qy	1393	ACCTCCAAGAACCAGGCCCGCTGATGCAAGATGGCAGATCTGTATACCCATTAGAGCCCC	1452
Db	469	-----GlnGlyProAlaGlyProGlyGlySerArgAspProGlyProProGly	484
Qy	1453	GAGAAATTCCTCTCTGGATCCCAGTTTGCAGCAACCCACACCCAGCTCACACAGCAA	1512
Db	485	AlaGlnGlyProAlaGlyProGlyGlySerArgAspProGlyProGlyYalaGlnGly	504
Qy	1513	AMCAATGGACAGGC-----CCAGAGGCTGAAGCAACACAGTGTCTC	1551
Db	505	ProAlaGlyProGlyGlySerArgAspProGlyProGlyYalaGlnGlyProAlaGly	524
Qy	1552	CCTTCTGGCTGTGTGGAGCCTCCCAAGTAACCACTATTATTTTACCTCTTTTCCCAAA	1611
Db	525	ProGlyGlySerArgAspProGlyPro-----	533
Qy	1612	CTGTGAGCATTTATGCTATGGCTGTGCAAGATCTGTTCAGTCCCTCT	1659
Db	534	ProGlyYalaHisGlyProAlaGlyProLysGlyYAlaHisGlyProAla	549

```

US-09-585-887-9
; Sequence 9, Application US/09585887
; Patent No. 6413742
; GENERAL INFORMATION:
; APPLICANT: Olsen, David R
; APPLICANT: Chang, Robert
; APPLICANT: McMullin, Hugh
; APPLICANT: Hitzeman, Ronald A.
; APPLICANT: Chisholm, George
; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
; TITLE OF INVENTION: CELLS
; FILE REFERENCE: 225002030400
; CURRENT APPLICATION NUMBER: US/09/585,887
; CURRENT FILING DATE: 2000-05-31
; PRIOR APPLICATION NUMBER: 09/289,578
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: 60/084,828
; PRIOR FILING DATE: 1998-05-08
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: patentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1461
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-585-887-9

```

Alignment Scores:						
Pred. No.:	9.03e-07	Length:	1461			
Score:	177.50	Matches:	163			
Percent Similarity:	27.59%	Conservative:	29			
Best Local Similarity:	23.42%	Mismatches:	222			
Query Match:	5.68%	Indels:	282			
DB:	4	Gaps:	33			
US-09-762-021A-1 (1-1710) x US-09-585-887-9 (1-1461)						
Qy 1 GCCAGAGCGACTGAAGACCACTGGTGC-----GAAGGCTCTGGAGGAAAGAGCTGGAGCA 54						

; APPLICANT: Hitzeman, Ronald A.
 ; APPLICANT: Chisholm, George
 ; TITLE OF INVENTION: NOVEL METHODS FOR THE PRODUCTION OF GELATIN AND
 ; TITLE OF INVENTION: FULL-LENGTH TRIPLE HELICAL COLLAGEN IN RECOMBINANT
 ; TITLE OF INVENTION: CELLS
 ; FILE REFERENCE: 225002030400
 ; CURRENT APPLICATION NUMBER: US/09/289,578
 ; CURRENT FILING DATE: 1999-04-10
 ; PRIOR APPLICATION NUMBER: 60/084,828
 ; PRIOR FILING DATE: 1998-05-08
 ; NUMBER OF SEQ ID NOS: 11
 ; SOFTWARE: PatentIn ver. 2.0
 ; SEQ ID NO 9
 ; LENGTH: 1461
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 ; US-09-289-578-9

Alignment Scores:

Pred. No.:	9,03e-07	Length:	1461
Score:	177.50	Matches:	163
Percent Similarity:	27.59%	Conservative:	29
Best Local Similarity:	23.42%	Mismatches:	222
Query Match:	5.68%	Indels:	282
DB:	4	Gaps:	33

US-09-762-021a-1 (1-1710) x US-09-289-578-9 (1-1461)
 QY 1 GCGAGCGGCTGAGACAGCCTGCA-----GAAGGCTCGGAGGAGAGCTGGAGCA 54
 DB 497 GlyValAlaGlyProLysGlyProAlaGlyGluArgGlySerProGlyProAlaGly--- 515
 QY 55 AAGACCTGACCTGGAGGCTTCACCGAGCGCCAGCAGATGGAGGGGCGCTGTATGA 114
 DB 516 -----ProLysGlySerProGly-----GluAlaGlyArgProGly 527
 QY 115 AAGCGCGCTCCCTATGGAGCGGACGCTATCTGGAGCC-----GGGATCCC 162
 DB 528 GluAlaGlyLeuProGlyAlaLysGlyLeuThrGlySerProGlySerProGlyProAsp 547
 QY 163 TCCAGACAGCCCGCCACCA-----GAGGACCTAGACACAGCCTCCACCATC 210
 DB 548 GlyLysThrGlyProGlyProAlaGlyGlnAspGlyArgProGlyProPro----- 565
 QY 211 CCCAAGGCGCTGCCAGCGCCACAGCTGCCGAGAACCAAGTGCTTTACTCTGCCTCC 270
 DB 565 ----- 565
 QY 271 TCCAAGGCGGTCTCTTCCCCGAGGAGCCAGAGAGGAGGAGGAGTGTGAA----- 324
 DB 566 -----GlyProGlyAlaArgGlyGlnAlaGlyValMet 577
 QY 325 -----CCATGTCCTAAGGACATTGACGTGTTTCATGGGAAAGCTGGAGAA----- 369
 DB 578 GlyPheProGlyProLysGlyAlaAlaGlyGluProGlyLysAlaGlyGluArgGlyVal 597
 QY 370 -----GGCCCA-----GGCAAGACACGACGAAGAATAATTTGG 405
 DB 598 ProGlyProProGlyAlaValGlyProAlaGlyLysAspGlyGlu----- 612
 QY 406 GAAAAAACAAGGACCGAGGAGGTCTCACCGAGCCACAGTACATTGACTGCTTCAGAA 465
 DB 613 AlaGlyAlaGlnGlyPro-----ProGly----- 620
 QY 466 GATCAAGTACAGCTCAACCTCTGGGAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGC 525
 DB 621 -----ProAlaGlyProAlaGlyGluArgGlyGlnGlyProAla 634
 QY 526 CCCTGAGCTCGTACACATCTCTCAAGTCCCTGAATCTCACTCTGGCCAGGTGCCCTGA 585
 DB 635 GlySerProGlyPheGlnGlyLeuProGlyProAlaGlyProProGlyGluAla----- 652
 QY 586 GCGTGGCCTPAGCAGC-----CCAAGTGATCTACCCCTCTCACCCCTCAAGACTAT 636

Db	653	GlyLysProGlyGluGlnGlyValProGlyAspLeuGlyAlaProGlyPro-----	669
QY	637	CAACCTGCTACAGTCTCTAAGCCACCTGAGAGTAACCTTTGGATGGG-----	687
Db	670	-----SerGlyAlaArgGlyGluArgGlyPheProGlyGluArgGly 683	
QY	688	---GTTGGGCCAGCCTGGACCACTAGCCGGCCGACCTGGACAGCGCATGAGCCCC-----	740
Db	684	ValGlnGlyPro--ProGlyPro---AlaGlyProArgGlyAlaAsnGlyAlaProGlyA 702	
QY	741	-----TCCCTACCACCCACATCTCTCAG-----	764
Db	702	snAspGlyAlaLysGlyAspAlaGlyAlaProGlyAlaProGlySerGlnGlyAlaProG 722	
QY	764	-----	764
Db	722	lyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeuProGlyProLysGlyAspA 742	
QY	765	-----A 765	
Db	742	rgGlyAspAlaGlyProLysGlyAlaAspGlySerProGlyLysAspGlyValArgGlyL 762	
QY	766	TCAGTGGCAACTTCCAGAGCCCT---CCAGCCACACCCCTTAGATACAGGACCCTGT 822	
Db	762	euThrGlyProIleGlyProGlyProAlaGlyAlaProGlyAspLysGlyGluSerG 782	
QY	823	TTCCCTTCGGC-----GGGGAAGTCATAGTTAGGAGGACCACTCACACTTTCCTCAGGA 876	
Db	782	lyProSerGlyProAlaGlyProThrGlyAlaArgGlyAlaProGly----- 797	
QY	877	GAAGACACACAACCATGATGACCTCGGAGCCCACTCACTCA----- 920	
Db	798	-----AspArgGlyGluProGlyProGlyProAlaGlyP 810	
QY	920	----- 920	
Db	810	heAlaGlyProProGlyAlaAspGlyGlnProGlyAlaLysGlyGluProGlyAspAlaG 830	
QY	921	-----GGCCCTCCAGCCCAACCTCCCGAGCCAGCCCTCAAAATGCA 963	
Db	830	lyAlaGlyGlyAspAlaGlyProGlyPro--AlaGlyProAlaGlyProProGly--- 848	
QY	964	AGTCTGTGACGAGTTGAAGCTAGGAACCCAGCGGAGCACTGCTGTGTCAGGGAGAGAA 1023	
Db	849	-----ProIleGlyAsnValGlyAlaProGlyAlaLys 859	
QY	1024	GCT-----GGAGGTTCTGGACCAAGCAAGCGGTG----- 1053	
Db	860	GlyAlaArgGlySerAlaGlyProGlyAlaThrGlyPheProGlyAlaAlaGlyArg 879	
QY	1054	-----GTGGCTGT 1062	
Db	880	ValGlyProProGlyProSerGlyAsnAlaGlyProProGlyProProGlyProAlaGly 899	
QY	1063	GAAGAATGAGCGGAGCGGAGCGG----- 1086	
Db	900	LysGluGlyGlyLysGlyProArgGlyGluThrGlyProAlaGlyArgProGlyGluVal 919	
QY	1087	-----CTACATCCACGAACATCTCTGGAGCCCT---ACA 1119	
Db	920	GlyProProGlyProProGlyProAlaGlyGluLysGlySerProGlyAlaAspGlyPro 939	
QY	1120	GCCGGGACCCCTGGACCCAGCGGCGAGTCCCTCGGTTCGAATCTTCGACTTAG 1179	
Db	940	AlaGlyAlaProGlyThrProGlyProGlnGlyIleAlaGlyGlnArgGlyValValGly 959	
QY	1180	CTCAGGCGCTGAAGAGTGCACAGCTGGCTGCAGCGACGAGAACTTCTCCACTGCCACGGT 1239	
Db	960	LeuProGlyGlnArgGlyGluArgGlyPheProGlyLeuPro----- 973	
QY	1240	GAGGACACTTGGGTCCCTGAGCGGGAGCGCAGCTACTTCGCATTAAGACCTGGGAGCTACA 1299	

Db 974 -----GlyProSerGlyGluProGlyLysGlnGlyProSer----GlyAlaSer 988
Qy 1300 GATGCTATGTCACAGGAGGCCACCAATCCTGCTCCGGCTGGAGCTCTCAGAAGGAT 1359
Db 989 GlyGluArgGlyProProGlyProMetGlyProGlyLeuAlaGlyProGlyGlu 1008
Qy 1360 GCTGGG---GATNAGCCCTTAGGACACAGCTTAGACACCTCCAGAACCAGGCCGCTG 1416
Db 1009 SerGlyArgGluGlyAlaProGlyAlaGlySerProGlyArgAspGlySerProGly 1028
Qy 1417 ATGCAAGATGCCAGATCTGATACCATAGAGCCCGGAGATTCTCTTGGATCCAG 1476
Db 1029 AlaLysGlyAspArgGlyGluThrGlyProAlaGlyProGlyAlaProGlyAlaPro 1048
Qy 1477 TTTCCAGCAAAACCCACACCCAGCTC-----ACACAGCAAAACAATGAGAGGCCCA 1530
Db 1049 ValAlaProGlyProValGlyProAlaGlyLysSerGlyAspArgGlyGluThrGlyPro 1068
Qy 1531 GAGCCTGAAGCAACAGTGTCCCTTCTGGCTGTGTGGAGCC 1572
Db 1069 AlaGlyProAlaGly-----ProValGlyProValGlyAla 1080

RESULT 14

US-08-642-255-73
; Sequence 73, Application US/08642255
; Patent No. 5773249
; GENERAL INFORMATION:
; APPLICANT: CAPPELLO, Joseph
; APPLICANT: FERRARI, Franco A.
; TITLE OF INVENTION: High Molecular Weight Collagen-Like
; TITLE OF INVENTION: Protein Polymers
; NUMBER OF SEQUENCES: 135
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FLEHR, HOEBACH, TEST, ALBRITTON & HERBERT
; STREET: 4 Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/642,255
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: ROWLAND, Bertram I.
; REGISTRATION NUMBER: 20,015
; REFERENCE/DOCKET NUMBER: A55556-3/BIR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 494-8700
; TELEFAX: (415) 494-8771
; TELEX: 910 277299 FHT UR
; INFORMATION FOR SEQ ID NO: 73:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 633 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-642-255-73

Alignment Scores:
Pred. No.: 7.92e-07
Score: 176.50
Percent Similarity: 29.25%
Best Local Similarity: 25.16%
Query Match: 5.64%
DB: 1
Length: 633
Matches: 160
Conservative: 26
Mismatch: 245
Indels: 205
Gaps: 36

US-09-762-021a-1 (1-1710) x US-08-642-255-73 (1-633)
Qy 25 GCAGAGGCTCTGGAGGAAGAGCTGGAGCAAGACCTCGACTTGGAGGCTTCAGCCAGG 84
Db 89 AlaGlnGlyProAlaGlyProGlyGlySerArgGlyAspProGlyProGlyAlaGln 108
Qy 85 CCAGGACAGATGAGGGGGCTCTGCTATGAAAGGCCCTCCCTATGAGCA----- 135
Db 109 GlyProAlaGlyProGlyGlySerArgGlyAspProGlyProGlyAlaGlnGlyPro 128
Qy 136 -----GGCAGCTATCTGGAGCGGGGATCCCTCCAGACAGCCCA----- 177
Db 129 AlaGlyProGlyGlySerArgGlyAspProGlyProGlyAlaGlnGlyProAlaGly 148
Qy 178 -----CCAGAGGACCTAGAGCAGAGCTCCACCATCCCAAG 216
Db 149 ProGlyGlySerArgGlyAspProGlyProGlyAlaHisGlyProAlaGlyProLys 168
Qy 217 GCCCTGCCAGCCACACAGTGCCTCCGAGAACCAAGTCTTACTGTCTCTCTCCAG 276
Db 169 GlyAlaHisGlyPro-----AlaGlyProLys 177
Qy 277 GCGTCTCTTCCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 336
Db 178 GlyAlaHisGlyProAlaGlyProLysGlyAlaHisGlyProAlaGly-----ProLys 195
Qy 337 GGACATTGAGCTGTTTCATGGAAAGCTGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 396
Db 196 Gly-----AlaGlnGlyProAlaGlyProGlyGlySerArgGlyAspProGlyPro 212
Qy 397 GAAATTGGGAAAAAACAAGGAGGAGG-----AGGTCTCACCCAGGACAGTACAT 450
Db 213 ProGlyAlaGlnGlyProAlaGlyProGlyGlySerArgGlyAspProGly----- 229
Qy 451 TGACTGCTTCCAGAGATCAAGTACAGCTTCAACCTCTCGGAGGAGGAGGAGGAGGAGGAG 507
Db 230 -----ProGlyAlaGlnGlyProAlaGly 238
Qy 508 -----GCTGAAGGAGAGAGTGCCTGAGCTCGTACATCTCTTCAAGTCCCT 558
Db 239 ProGlyGlySerArgGlyAspProGlyPro----- 248
Qy 559 GAATTCATCTGGC---CAGGTGCCCTGAGGTGGCTAGCAGCCCAAGTATCTCACC 615
Db 249 -----ProGlyAlaGlnGlyProAlaGlyProGlyGlySerArgGlyAspProGly 265
Qy 616 CCTCCTCACCCCTAAAGCTATCAACCTGCTACAGTCTCTGCTAGGCCACCTGAGAGTAA 675
Db 266 ProGly-----AlaGlnGlyProAla--GlyPro----- 275
Qy 676 CTTTGGATGGGTGGGCCCGAGCTGAGCA-----CTAGCCGGCGCA 720
Db 276 -----GlyGlySerArgGlyAspProGlyProGlyProGlyAlaGlnGlyProAlaGlyProG 294
Qy 721 CTGACAGGCGATGAGCCCTGCGCTACCAACCCACATCTCAGATGACTGCAACTTC 780
Db 294 LyGlySerArgGlyAspProGlyProGlyAlaHisGly----- 307
Qy 781 AGAGCCCTCCAGCAAGCACCCTTAGGATACAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 840
Db 308 -----ProAla--GlyProLysGlyAlaHisGlyProAlaGlyProLys----- 321
Qy 841 TCATAGTTAGGAGGAGCACCCTCACACTTCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAG 900
Db 322 -----GlyAlaHisGlyProAlaGlyProLysGlyAlaHisGlyProAl 336
Qy 901 G-----CCTGGGAGCC-----AACTCCAGGCC 924
Db 336 aGlyProLysGlyAlaGlnGlyProAlaGlyProGlyGlySerArgGlyAsp-ProGlyP 356
Qy 925 CTCAGCCCCAACCTGCGCCAGCAGCCCTGAAATGCAAGTCTTGTACAGTTTGAAGC 984
Db 356 roProGlyAlaGlnGlyProAlaGlyProGlyGlySerArgGly-----Asp 372


```
Db      239 roAlaGlyProProGlyAlaProGlyProAlaGlyPro----- 251
Qy      778 TCCAGAGCCCTCCAGCCCAAGCACCTTAGGATACCAGGACCTGTTCCCTTCGGCGGGG 837
Db      252 -----ProGlyAlaProGlyPro----- 257
Qy      838 AAGTCATAGGTTAGGAGCAGCTCACACTTTCCTCAGGAGAAGACACACAACCATGACCC 897
Db      258 -----AlaGlyProPro----- 261
Qy      898 TCAGCCTGGGGACCCCAACTCCA---GGCCCTCCAGCCCCCAACCTGCCAGCCAGCCCT 954
Db      262 -----GlyAlaProGlyProAlaGlyProProGlyAlaHis-GlyProAlaGlyPro 278
Qy      955 GAAATGCAAGCTCTGTACGAGTTTGAAGCTAGGAAACCCAGCGGAACCTGCTGTGTCCTCA 1014
Db      279 LysGlyAlaHis-----GlyPro 284
Qy      1015 GGGAGAGAAGCTGGAGGTTCTGGACCACAGCAGCGGTGGTGGCTGGAAGAAATGAGGC 1074
Db      285 AlaGlyProLysGlyAlaHisGlyPro-----Ala 294
Qy      1075 GGGAGCGAGCGGTACATTCACAGCAACATCCT---GGAGCCCT---ACAGCCGGGAC 1128
Db      295 GlyProLysGlyAlaHisGlyProAlaGlyProLysGlyAlaProGlyProAlaGlyPro 314
Qy      1129 CCCTGGGACCAGCGCCAGTCACCTCTCGGTTCCAAATGCTTCGACTTAGCTCGAGGCC 1188
Db      315 ProGlyAlaProGlyProAlaGlyProProGlyAlaProGly----- 328
Qy      1189 TGAAGAGTCAAGACTGGCTGCAGGAGAGAACTTCTCCACTGCCACGGTGAGGACACT 1248
Db      329 -----ProAlaGly----- 331
Qy      1249 TGGTCCCTGACGGGAGCCACTTCCGATAAGACTTGGGAGCTACAGATGCTATG 1308
Db      332 -----ProProGlyAlaProGlyProAla-----GlyProProGlyAlaPro 345
Qy      1309 TCCACAGGAGCCCGCAGCAAT---CCTGTCCCGCTGA-----GGCTGCAGAAGGAT 1359
Db      346 GlyProAlaGlyProProGlyAlaProGlyProAlaGlyProProGlyAlaProGlyPro 365
Qy      1360 GCTGGGATAAGCCCTTAGGCACCAGCTTAGACACCTCCAAGAACCCAGGCCCGCTCATG 1419
Db      366 AlaGly-----ProProGlyAlaProGlyProAlaGlyProProGlyAlaPro----- 381
Qy      1420 CAAGATGGCAGATCTGATACCATTAGAGCCCGAGAATTCTCTCTGGATCCAGTTT 1479
Db      382 -----GlyProAlaGlyProProGly 388
Qy      1480 GCAGCAAAACCCACCCAGCTCACAGCAAAACAAATGGACAGGCCCGAGGCTGAA 1539
Db      389 AlaProGlyProAlaGlyPro-----ProGlyAlaPro 399
Qy      1540 GCAAAACAGTGTCCCTTCTGGCTGTGTTGA---GCCCTCCAGTAACCACTATTATTT 1596
Db      400 GlyProAlaGlyProProGlyAlaProGlyProAlaGlyPro----- 413
Qy      1597 TACCTCTTCCAAACCTGGAGCATTTATGCTAGGCTTGTCAAGAATCTGTTTCAGTCCC 1656
Db      414 -----ProGlyAlaHisGlyProAlaGlyProLysGlyAlaHisGlyPro 428
Qy      1657 TCTCCTTCTCAATAAAGCAT 1677
Db      429 AlaGlyProLysGlyAlaHis 435
```

Search completed: February 25, 2003, 19:45:22
Job time : 42.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 05:25:51 ; Search time 3042 Seconds
(without alignments)
16359.565 Million cell updates/sec

Title: US-09-762-021a-1

Perfect score: 1710

Sequence: 1 ggcagagcactgaagacca.....aaaaaaaaaataangataaa 1710

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sts.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pin.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_sy.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1696.6	99.2	2220	6	AX058186 Sequence
2	1683.6	98.5	2255	9	BC012926 Homo sapi
3	1508.6	88.2	2148	9	AK025175 Homo sapi
4	1424.4	83.3	1973	9	AY074930 Homo sapi
5	1334.6	78.0	2159	6	AX322735 Sequence
6	918	53.7	2275	10	BC014734 Mus muscu
7	913.2	53.4	2329	10	AY074932 Mus muscu
8	389	22.7	401	6	AX407643 Sequence
9	324.4	19.0	133451	9	AL158847 Human DNA
10	322.8	18.9	162480	2	AC092860 Pan trogl
11	218.4	12.8	224	11	G25672 human STS E
12	218.4	12.8	224	11	G41866 SHGC-32779
13	155.6	9.1	197	6	AR044148 Sequence
14	117.6	6.9	113285	10	AC018461 Mus muscu
15	117.6	6.9	198459	2	AC079042 Mus muscu
16	110.8	6.5	185397	2	AC095838 Rattus no
17	102.6	6.0	185397	2	AC095838 Rattus no
18	91.8	5.4	2527	9	AY074929 Homo sapi
19	91.8	5.4	2646	9	BC002474 Homo sapi
20	91.8	5.4	3074	9	AK025824 Homo sapi
21	91.8	5.4	3102	9	AK025588 Homo sapi
22	88	5.1	2557	9	AY074928 Homo sapi
23	86.4	5.1	2201	9	BC015763 Homo sapi
24	86.4	5.1	2228	9	AK000265 Homo sapi
25	86.4	5.1	2685	6	AX399900 Sequence
26	85	5.0	1829	9	AF282167 Homo sapi
27	84.8	5.0	2169	9	BC004907 Homo sapi
28	83.2	4.9	125020	9	AF429315 Homo sapi
29	79.8	4.7	3079	10	BC005492 Mus muscu
30	70.4	4.1	2151	10	AY074931 Mus muscu
31	70.4	4.1	2995	10	BC027043 Mus muscu
32	66	3.9	2233	9	AK094539 Homo sapi
33	64.2	3.8	125020	9	AF429315 Homo sapi
34	63.8	3.7	1827	3	AY075572 Drosophil
35	63.8	3.7	3035	3	AF208262 Drosophil
36	62.6	3.7	90816	2	AC017493 Drosophil
37	62.6	3.7	102389	3	AC004546 Drosophil
38	62.6	3.7	165652	3	AC092222 Drosophil
39	62.6	3.7	299378	3	AE003588 Drosophil
40	56.8	3.3	1977	9	AK057052 Homo sapi
41	56.4	3.3	1947	9	AK027765 Homo sapi
42	56.2	3.3	3365	10	BC009098 Mus muscu
43	55	3.2	514	6	AX321026 Sequence
44	55	3.2	3832	9	HSU12535 Human epide
45	55	3.2	3892	9	BC030010 Homo sapi

ALIGNMENTS

RESULT 1

AX058186

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

AX058186 Sequence 56 from Patent WO0077040. 2220 bp DNA linear PAT 17-JAN-2001

AX058186.1 GI:12310687

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2220)

Yue.H., Tang.Y.T., Hillman,J.L., Lal,P., Bandman,O., Baughn,M.R., Azimzai,Y., Yang,J., Reddy,R. and Lu,D.A.

Human intracellular signaling molecules

JOURNAL Patent: WO 0077040-A 56 21-DEC-2000;

Incyte Genomics, Inc. (US)

FEATURES Location/Qualifiers

source 1..2220

/organism="Homo sapiens"

/db_xref="taxon:9606"

BASE COUNT 549 a 696 c 593 g 382 t

ORIGIN

Query Match 99.2%; Score 1696.6; DB 6; Length 2220;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 1699; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 1 GCGAGAGCGACTGAAGACGACGCTCGAAGAGGCTCTGGAGGAAGAGCTGGAGCAAAAGACC 60
Db 518 GCGAGAGCGACTGAAGACGACGCTCGAAGAGGCTCTGGAGGAAGAGCTGGAGCAAAAGACC 577
Qy 61 TCGAGCTTGGAGGCTTACAGCAGGCGCAGGACAGATGAGAGGGGCTGTATGGAAGGCC 120
Db 578 TCGAGCTTGGAGGCTTACAGCAGGCGCAGGACAGATGAGAGGGGCTGTATGGAAGGCC 637
Qy 121 GCTCCCTATGAGGACGACGCTATCTGAGCGCGGGATCCCTCCAGAACAGCCCCACCA 180
Db 638 GCTCCCTATGAGGACGACGCTATCTGAGCGCGGGATCCCTCCAGAACAGCCCCACCA 697
Qy 181 GAGGACCTTAGAGCACAGCCTCCACCATCCCAAGGCCCTTGCACGCCACACCAAGTGC 240
Db 698 GAGGACCTTAGAGCACAGCCTCCACCATCCCAAGGCCCTTGCACGCCACACCAAGTGC 757
Qy 241 CCGAGAACCAAGTGGCTTACTCTGCTCCTCCCAAGGGGGTCTTCCCGCCGAGGAGCC 300
Db 758 CCGAGAACCAAGTGGCTTACTCTGCTCCTCCCAAGGGGGTCTTCCCGCCGAGGAGCC 817
Qy 301 AGAGAGGACGAGGAAGTGTGAACATGTCTTAAGGGACATTTGAGCTGTTCATGGGAAA 360
Db 818 AGAGAGGACGAGGAAGTGTGAACATGTCTTAAGGGACATTTGAGCTGTTCATGGGAAA 877
Qy 361 GCTGGAGAAGGCCCAAGCAAGACGAGGAGGAAGAAATTTGGGAAAAAACAAGGA 420
Db 878 GCTGGAGAAGGCCCAAGCAAGACGAGGAGGAAGAAATTTGGGAAAAAACAAGGA 937
Qy 421 CCAGGAGGTCTACCCAGGACAGTACATTCAGTCTTCCAGAGATCAAGTACAGTTC 480
Db 938 CCAGGAGGTCTACCCAGGACAGTACATTCAGTCTTCCAGAGATCAAGTACAGTTC 997
Qy 481 CAACCTCTGGAGGCTGGCCACCTGGCTGAAGGAGACAAGTGCCTGAGCTCGTACA 540
Db 998 CAACCTCTGGAGGCTGGCCACCTGGCTGAAGGAGACAAGTGCCTGAGCTCGTACA 1057
Qy 541 CATCCTCTTCAAGTCCCTGAACCTTCATCTGCGCCAGGTGCCCTGAGGCTGCGCTAGCAGC 600
Db 1058 CATCCTCTTCAAGTCCCTGAACCTTCATCTGCGCCAGGTGCCCTGAGGCTGCGCTAGCAGC 1117
Qy 601 CCAAGTATCTACCCCTCTCAGCCCTTAAAGCTATCAACCTGCTACAGTCTGTCFPAAG 660
Db 1118 CCAAGTATCTACCCCTCTCAGCCCTTAAAGCTATCAACCTGCTACAGTCTGTCFPAAG 1177
Qy 661 CCCACCTCAGAGTAACCTTTGGATGGGCTGGGCCAGCTGGACCACTAGCCGGGCCGA 720
Db 1178 CCCACCTCAGAGTAACCTTTGGATGGGCTGGGCCAGCTGGACCACTAGCCGGGCCGA 1237
Qy 721 CTGGACAGCGATGAGCCCTGCCCTACCAACCCACATTTCTCAGATGACTGGCAACTTCC 780
Db 1238 CTGGACAGCGATGAGCCCTGCCCTACCAACCCACATTTCTCGGATGACTGGCAACTTCC 1297
Qy 781 AGAGCCCTCCAGCAAGCACCTTTAGGATACAGGACCTGTTCCTTCCCTTCCGCGGGGAAG 840
Db 1298 AGAGCCCTCCAGCAAGCACCTTTAGGATACAGGACCTGTTCCTTCCCTTCCGCGGGGAAG 1357
Qy 841 TCATAGGTTAGGAGCACTCACACTTTCCTCAGGAGAGACACACACCAACCATGACCTCA 900
Db 1358 TCATAGGTTAGGAGCACTCACACTTTCCTCAGGAGAGACACACACCAACCATGACCTCA 1417

Qy 901 GCCTGGGAGCCCCAACATCCAGGCCCTCCAGCCCCAAACCTCCCGAGCAGCCCTGAAAAAT 960
Db 1418 GCCTGGGAGCCCCAACATCCAGGCCCTCCAGCCCCAAACCTCCCGAGCAGCCCTGAAAAAT 1477
Qy 961 GCAAGTCTTTACAGATTTGAAGCTTAGGAACCCACGCGAACTGACTGTGTTCCAGGGAGA 1020
Db 1478 GCAAGTCTTTACAGATTTGAAGCTTAGGAACCCACGCGAACTGACTGTGTTCCAGGGAGA 1537
Qy 1021 GAAGCTGGAGGTTCTGGACACACAGCGGTGGTGGCTGGTGAAGAATGAGCGGGGACG 1080
Db 1538 GAAGCTGGAGGTTCTGGACACACAGCGGTGGTGGCTGGTGAAGAATGAGCGGGGACG 1597
Qy 1081 GAGCGGCTACATTCGAAGCAACATCTCTGAGCCCTACAGCGGGGAGCCCTGGGAGCCCA 1140
Db 1598 GAGCGGCTACATTCGAAGCAACATCTCTGAGCCCTACAGCGGGGAGCCCTGGGAGCCCA 1657
Qy 1141 GGGCCAGTCAACCTCTCGGGTTCCAATGCTTCGACTTTCGACTTTCGAGGCTTGAAGAGTCCAC 1200
Db 1658 GGGCCAGTCAACCTCTCGGGTTCCAATGCTTCGACTTTCGAGGCTTGAAGAGTCCAC 1717
Qy 1201 AGACTGGCTGAGGACAGAACTTCTCCACTGCCACGCTGAGGACACTTGGTCCCTGAC 1260
Db 1718 AGACTGGCTGAGGACAGAACTTCTCCACTGCCACGCTGAGGACACTTGGTCCCTGAC 1777
Qy 1261 GGGAGGCGAGTACTTTCGCATAAGACCTGGGAGCTACAGATGCTATGTCACAGAGGCG 1320
Db 1778 GGGAGGCGAGTACTTTCGCATAAGACCTGGGAGCTACAGATGCTATGTCACAGAGGCG 1837
Qy 1321 CCCAGGATCTGTCGGGCTGGAGGCTGTCAGAAGGATGCTGGGATAGCCCTTAGGC 1380
Db 1838 CCCAGGATCTGTCGGGCTGGAGGCTGTCAGAAGGATGCTGGGATAGCCCTTAGGC 1897
Qy 1381 ACCAGCTTAGACACTTCCAGAACAGCCCGCTGATGCAAGATGGCAGATCTGATACC 1440
Db 1898 ACCAGCTTAGACACTTCCAGAACAGCCCGCTGATGCAAGATGGCAGATCTGATACC 1957
Qy 1441 CATTAGAGCCCGAGAAATCTCTTCTGGATCCAGTTCCTGAGCAAAACCCACACCCAG 1500
Db 1958 CATTAGAGCCCGAGAAATCTCTTCTGGATCCAGTTCCTGAGCAAAACCCACACCCAG 2017
Qy 1501 CTCACAGCAAAACAAATGACAGGCCAGGCTGAAGCAACAGTGCCTTCTGCG 1560
Db 2018 CTCACAGCAAAACAAATGACAGGCCAGGCTGAAGCAACAGTGCCTTCTGCG 2077
Qy 1561 TGTGTGGAGGCTCCCGAGTAAACCACTTATTTATTTTACCTCTTTCCCAACCTGGAGCA 1620
Db 2078 TGTGTGGAGGCTCCCGAGTAAACCACTTATTTATTTTACCTCTTTCCCAACCTGGAGCA 2137
Qy 1621 TTTATGCGCTAGGCTGTCAAGAATCTGTTGAGTCCCTCTCCTCTCAATAAAGCATCTT 1680
Db 2138 TTTATGCGCTAGGCTGTCAAGAATCTGTTGAGTCCCTCTCCTCTCAATAAAGCATCTT 2197
Qy 1681 CAAGCTTGTAAAAAATAAA 1703
Db 2198 CAAGCTTGTAAAAAATAAA 2220

RESULT 2

BC012926

LOCUS

DEFINITION

BC012926

ACCESSION

BC012926

VERSION

BC012926.1

KEYWORDS

MGC.

SOURCE

ORGANISM

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

1 (bases 1 to 2255)

STRAUSBERG, R.

Direct Submission

TITLE

JOURNAL

Submitted (20-AUG-2001) National Institutes of Health, Mammalian

BC012926 2255 bp mRNA linear PRI 22-AUG-2001
Homo sapiens, Similar to hypothetical protein FLJ21522, clone
MGC:16817 IMAGE:3853503, mRNA, complete cds.

BC012926

BC012926.1 GI:15277845

MGC.

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 2255)

STRAUSBERG, R.

Direct Submission

TITLE

JOURNAL

Submitted (20-AUG-2001) National Institutes of Health, Mammalian

Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgabbs@mail.nih.gov
Tissue Procurement: ARCC
cDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: villalon@bcm.tmc.edu
Villalon, D.K., Luna, R.A., Hale, S.M., Hulyk, S., Lu, X., Garcia, A.M., Holloway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W., Muzny, D.M., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 13 Row: a Column: 21
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 10437637.

FEATURES

source

Location/Qualifiers
1..2255
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="MGC:16817 IMAGE:3853503"
/tissue_type="Colon, adenocarcinoma"
/lab_host="NIH_MGC_65"
/note="Vector: pCMV-SPORT6"
146..1930
/codon_start=1
/product="Similar to hypothetical protein FLJ21522"
/protein_id="AAH12926.1"
/db_xref="GI:15277846"

CDS

/translation="MSRPSRAIYLHRKEYSONLTSEPTLLQHRVHLMTCQGSORV
QGPEDALOKLFMDAQGRVWSODLILQVRDGLQLLDIETKEELSYRLDSIQANVA
LNTSVNSILSTVQEPGLPGTSTLLFQCEVGAERLKTSLQALEEELQSRPLGG
LQGDWRGPMERPLPEQARYLEPIPPQPHQRTLEHSLPPSRPLRHTSARE
PSAFTLPPRRSSPEDEEVNLVRLHLEFPMGLKLEAKQAKTSRKKKFGKKNK
OQGLTQAOI IDQFIKYFNLLGLATWLKETSAPLWHILFKLSINFLARCPBAGL
AAQVLSPLTLKALINLLQSLSPSPSNLMMGLPAWTTSRADWTGDEPLPYPTFSD
WOLPESSQAPLGYDDPSLRGRSLGTSHPKPOKTHNHDPQPDNSRSPSPKPA
OPALKMVLVEFARNPRELTVQGEKLVDHPSKRWMLVKNEAGRSYIPSNILEPL
QPGTCTQCSRPVPMRLSRPEVTDWLQAEFNSTATVTLGSLTGSLSQLIRPG
ELQMLCPQAPRILSRLEAVRMGLSP"

BASE COUNT
ORIGIN

Query Match 98.5%; Score 1683.6; DB 9; Length 2255;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1699; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY	1	GGCAGAGCGACTGAAGACACCGCTGCAGAGGCTCTGGAGAGAGCTGGAGCAA---AG	57
DB	550	GGCAGAGCGACTGAAGACACCGCTGCAGAGGCTCTGGAGAGAGCTGGAGCAAAGCAG	609
QY	58	ACCTCGACTTGGAGGCTTCAGGACGCCAGGACAGATGGAGGGGCGCTGTATGAAAG	117
DB	610	ACCTCGACTTGGAGGCTTCAGGACGCCAGGACAGATGGAGGGGCGCTGTATGAAAG	669
QY	118	GCCGCTCCCTATGAGCAGCAGCGCTATCTGGAGCGGGATCCCTCCAGAACAGCCCCA	177
DB	670	GCCGCTCCCTATGAGCAGCAGCGCTATCTGGAGCGGGATCCCTCCAGAACAGCCCCA	729
QY	178	CCAGAGACCCCTAGAGCACAGCCTCCACCATCCCAAGGCCCTGCCAGGCCACACCCAG	237
DB	730	CCAGAGACCCCTAGAGCACAGCCTCCACCATCCCAAGGCCCTGCCAGGCCACACCCAG	789
QY	238	TGCCCGAGAACCAAGTGCCCTTTACTTGCCCTCTCCAAAGCGGTCTCTTCCGCCGAGGA	297

DB	790	TGCCCCGAGAACCAAGTGCCCTTTACTTGCCCTCTCCAAAGCGGTCTCTTCCGCCGAGGA	849
QY	298	CCCAGAGAGGAGCAGGAGAGTGTCTGAACCATGTCTTAAGGGACATTTAGCTGTTCATGGG	357
DB	850	CCCAGAGAGGAGCAGGAGAGTGTCTGAACCATGTCTTAAGGGACATTTAGCTGTTCATGGG	909
QY	358	AAAGCTGGAGAGGCGCCAGGCCAAAGACCAAGACAGCAAGAGAGAGAAATTTGGGAAAAAACA	417
DB	910	AAAGCTGGAGAGGCGCCAGGCCAAAGACCAAGACAGCAAGAGAGAGAAATTTGGGAAAAAACA	969
QY	418	GGACACGAGGAGGTCTCACCCAGGCACAGTACATTTGACTGCTTCCAGAGATCAAGTACAG	477
DB	970	GGACACGAGGAGGTCTCACCCAGGCACAGTACATTTGACTGCTTCCAGAGATCAAGTACAG	1029
QY	478	CTTCAACCTCTCTGGGAAAGGCTGGCCACCTTGGCTGGAAGGAGACAAAGTGCCTCAGCTCGT	537
DB	1030	CTTCAACCTCTCTGGGAAAGGCTGGCCACCTTGGCTGGAAGGAGACAAAGTGCCTCAGCTCGT	1089
QY	538	ACACATCTCTTCAAGTTCCTTGAACCTTCATTCCTTGGCCAGGTGCCCTTGAGCTGGCCTAGC	597
DB	1090	ACACATCTCTTCAAGTTCCTTGAACCTTCATTCCTTGGCCAGGTGCCCTTGAGCTGGCCTAGC	1149
QY	598	AGCCCAAGTGATCTCACCCCTCTCTCACCCCTTAAAGCTATCAACCTGCTACAGTCTCTGTCT	657
DB	1150	AGCCCAAGTGATCTCACCCCTCTCTCACCCCTTAAAGCTATCAACCTGCTACAGTCTCTGTCT	1209
QY	658	AAGCCACCTTGACAGTAACCTTTTGGATGGGTTGGCCAGCTGGACCACTAGCCGGGC	717
DB	1210	AAGCCACCTTGACAGTAACCTTTTGGATGGGTTGGCCAGCTGGACCACTAGCCGGGC	1269
QY	718	CGACTGGACAGGCGATGAGCCCTGCTCCCTACCAACCCACATTTCTCAGATGACTTGGCAACT	777
DB	1270	CGACTGGACAGGCGATGAGCCCTGCTCCCTACCAACCCACATTTCTCAGATGACTTGGCAACT	1329
QY	778	TCAGAGCCCTTCAGCCAGCAGCCCTTAGATACAGGACCCCTTGTTCCTTCGGGGGG	837
DB	1330	TCAGAGCCCTTCAGCCAGCAGCCCTTAGATACAGGACCCCTTGTTCCTTCGGGGGG	1389
QY	838	AGCTCATAGTTAGGAGCAGCTCACACTTTCTCAGGAGAACACACAAACCATGACCC	897
DB	1390	AGCTCATAGTTAGGAGCAGCTCACACTTTCTCAGGAGAACACACAAACCATGACCC	1449
QY	898	TACAGCTGGGAGCCCAACTTCCAGGCGCTCCAGCCCAACCTGCCAGCAGCCCTGAA	957
DB	1450	TACAGCTGGGAGCCCAACTTCCAGGCGCTCCAGCCCAACCTGCCAGCAGCCCTGAA	1509
QY	958	AATGCAAGCTTGTACAGTTTGAAGTAGGAGACCCACGGGAACCTGCTGTGTGAGGG	1017
DB	1510	AATGCAAGCTTGTACAGTTTGAAGTAGGAGACCCACGGGAACCTGCTGTGTGAGGG	1569
QY	1018	AGAGAAGCTGGAGGTTCTGGACCAAGAGCGGTGGTGGCTGTGAAGAAATCAGCGGG	1077
DB	1570	AGAGAAGCTGGAGGTTCTGGACCAAGAGCGGTGGTGGCTGTGAAGAAATCAGCGGG	1629
QY	1078	ACGAGCGGCTACATTTCCAAAGCAACATCTTGGAGCCCTTACAGCCGGGGACCCCTGGAC	1137
DB	1630	ACGAGCGGCTACATTTCCAAAGCAACATCTTGGAGCCCTTACAGCCGGGGACCCCTGGAC	1689
QY	1138	CCAGGCGCAGTACCCCTCTCGGGTTCCAATGCTTCACACTAGCTCAGGCGCTGAAAGGT	1197
DB	1690	CCAGGCGCAGTACCCCTCTCGGGTTCCAATGCTTCACACTAGCTCAGGCGCTGAAAGGT	1749
QY	1198	CACAGACTGGCTCCAGGAGAGAACTTCTCCACTGCCAGGTGAGGACACTTGGGTCCCT	1257
DB	1750	CACAGACTGGCTCCAGGAGAGAACTTCTCCACTGCCAGGTGAGGACACTTGGGTCCCT	1809
QY	1258	GACGGGAGCCAGCTACTTTCGCAATAAGACTTGGGAGCTACAGATGCTATGTCCACAGGA	1317
DB	1810	GACGGGAGCCAGCTACTTTCGCAATAAGACTTGGGAGCTACAGATGCTATGTCCACAGGA	1869
QY	1318	GGCCCCACGAATCTGTCCGGGCTGGAGGCTGTCAAGAGATCTGGGGATAGCCCTTA	1377
DB	1870	GGCCCCACGAATCTGTCCGGGCTGGAGGCTGTCAAGAGATCTGGGGATAGCCCTTA	1929

Qy 1378 GGCACAGCTTAGACACCTCCAAAGAACAGCCGCCGCTGATGCAAGATGGCAGATCTGAT 1437
D 1378 GGCACAGCTTAGACACCTCCAAAGAACAGCCGCCGCTGATGCAAGATGGCAGATCTGAT 1989
Qy 1438 ACCAATTAGACCCCGGAGAAATCTCTTCTGGATCCAGTTTGGAGCAAAACCCACACCC 1497
D 1438 ACCAATTAGACCCCGGAGAAATCTCTTCTGGATCCAGTTTGGAGCAAAACCCACACCC 2049
Qy 1498 CAGCTCACAGCAAAACAAATGACAGCCGACAGGCTGAGCAAAACAGTCTCCCTTCT 1557
D 1498 CAGCTCACAGCAAAACAAATGACAGCCGACAGGCTGAGCAAAACAGTCTCCCTTCT 2109
Qy 1558 GGCTGTGTGGAGCTCCCGAGTACACCACTATTATTTTACCTCTTTCCCAACCTGGA 1617
D 1558 GGCTGTGTGGAGCTCCCGAGTACACCACTATTATTTTACCTCTTTCCCAACCTGGA 2169
Qy 1618 GCATTTATGCTAGCTTGTGCAAGATCTGTCAGTCCCTCTCTCTCAATAAAGCAT 1677
D 1618 GCATTTATGCTAGCTTGTGCAAGATCTGTCAGTCCCTCTCTCTCAATAAAGCAT 2229
Qy 1678 CTTCAGCTCTGAAAAAATAA 1703
D 1678 CTTCAGCTCTGAAAAAATAA 2255
RESULT 3
AK025175 2148 bp mRNA linear PRI 29-SEP-2000
LOCUS Homo sapiens cDNA: FLJ21522 fis, clone COL05884.
DEFINITION AK025175
ACCESSION AK025175.1 GI:10437637
VERSION oligo capping: fis (full insert sequence).
KEYWORDS Homo sapiens colon cDNA to mRNA, clone_lib:COL
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Kawabata, A., Hikiji, T., Kobatake, N., Inagaki, H., Ikema, Y.,
Okamoto, S., Okitani, R., Ota, T., Suzuki, Y., Obayashi, M., Nishi, T.,
Shibahara, T., Tanaka, T., Nakamura, Y., Isogai, T. and Sugano, S.
NEDO human cDNA sequencing project
Unpublished
2 (bases 1 to 2148)
Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,
Shibahara, T., Tanaka, T. and Nakamura, Y.
Direct Submission
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing; Research Association for Biotechnology; cDNA library
construction, 5'- & 3'-end one pass sequencing; Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source
1. .2148
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="COL05884"
/tissue_type="colon"
/clone_lib="COL"
/note="cloning vector pME18SFL3"
128. .1819
/note="unnamed protein product"
/codon_start=1
/protein_id="BAB15081.1"
/db_xref="GI:10437638"
/translation="MSRPSSRAIYLRKREYSQNLTSPTLLQHRVHMLTCKQGSQRV

QGPEDALQKLFEMDAQGRVWSQDILQVRDGLQLDIETKEEUDSYRLDSIQAMNVA
LNTQSNLSITVPGTSTLLFQCOEVAERLKTSLQKALKEEELQORRLGGL
QPODRMRGPAMERPLPMEQARLEPQPHQRTLEHSLPSPRPPLRHPSAREP
SAFTLPPRRSSPEDPERDEELNHWLRDIELFMGLKKAQKTSRKKFKGKKKQKQ
GGLTOAYIDCFOKIKYSPNLGLRLATWLKETSAPLELVHILFKSLNFIARCP
AQVISELLPFAINLLOSLSPRESNLWMLGPAWTTSRADWTGDEPLPYQPTFSDW
QLPEPSSQAPLGYQDVPVSLRSPSPPAQPAKMGVLYEFARNPRELTVVQGGKLEVL
DHKRWMLVKNRAGSGYIPSNILEPLQPGTQGSRSRVPMLRSLSSRPBEVTDHL
QAEFNFSATVKTGLSGLTGSQRLIRPGEQMLCPQEPAPRLSLRLEAVRRMLGISP"
BASE COUNT 533 a 669 c 576 g 370 t
ORIGIN
Query Match 88.2%; Score 1508.6; DB 9; Length 2148;
Best Local Similarity 94.4%; Pred. No. 0;
Matches 1611; Conservative 0; Mismatches 5; Indels 90; Gaps 1;
Qy 1 GGCAGAGGACTGAAGACAGCCCTGCAGAGGCTCTGGAGAGAGCTGGAGCAAGACC 60
D 532 GGCAGAGGACTGAAGACAGCCCTGCAGAGGCTCTGGAGAGAGCTGGAGCAAGACC 591
Qy 61 TCGACTTTGGAGGCTTCAGCCAGGAGGACAGATGAGAGGGGCTGCTATGGAAAGGCC 120
D 592 TCGACTTTGGAGGCTTCAGCCAGGAGGACAGATGAGAGGGGCTGCTATGGAAAGGCC 651
Qy 121 GCTCCCTATGAGGAGCAGCAGCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCCACCA 180
D 652 GCTCCCTATGAGGAGCAGCAGCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCCACCA 711
Qy 181 GAGACCCCTAGAGCAGCAGCCCTCCCAACCATCCCCAAGCCCTGCCAGCCACACAGTGC 240
D 712 GAGACCCCTAGAGCAGCAGCCCTCCCAACCATCCCCAAGCCCTGCCAGCCACACAGTGC 771
Qy 241 CCGAGAACCAGTGCCTTTACTCTGCCTCTCCCAAGGGCGCTCTTCCCCCGAGGAGGCC 300
D 772 CCGAGAACCAGTGCCTTTACTCTGCCTCTCCCAAGGGCGCTCTTCCCCCGAGGAGGCC 831
Qy 301 AGAGAGGAGCAGGAGGAGTGTGAACCATGTCTTAAAGGACATTTGAGTGTTCATGGGAAA 360
D 832 AGAGAGGAGCAGGAGGAGTGTGAACCATGTCTTAAAGGACATTTGAGTGTTCATGGGAAA 891
Qy 361 GCTGAGAGGCCCCAGCAAGACAGCAGCAAGAGAAATTTGGGAAAAAACACAGGA 420
D 892 GCTGAGAGGCCCCAGCAAGACAGCAGCAAGAGAAATTTGGGAAAAAACACAGGA 951
Qy 421 CCAGGAGGCTCTACCCAGGACACAGTACATTTGACTTCTCCAGAGATCAAGTACAGCTT 480
D 952 CCAGGAGGCTCTACCCAGGACACAGTACATTTGACTTCTCCAGAGATCAAGTACAGCTT 1011
Qy 481 CAACCTCTGGAAGGCTGGCCACCTGGCTGAAGAGACAAAGTCCCTGAGCTGCTGTAAG 540
D 1012 CAACCTCTGGAAGGCTGGCCACCTGGCTGAAGAGACAAAGTCCCTGAGCTGCTGTAAG 1071
Qy 541 CATCTCTTCAAGTCCCTGAACTTTCATCTCGCCAGGCTGCCCTGAGGCTGGCTAGCAGC 600
D 1072 CATCTCTTCAAGTCCCTGAACTTTCATCTCGCCAGGCTGCCCTGAGGCTGGCTAGCAGC 1131
Qy 601 CCAAGTGTATCTACCCCTCTCCACCCCTTAAAGCTATCAACCTGCTTACAGTCTCTGTAAG 660
D 1132 CCAAGTGTATCTACCCCTCTCCACCCCTTAAAGCTATCAACCTGCTTACAGTCTCTGTAAG 1191
Qy 661 CCCACCTGAGAGTAACTTTTGGATGGGGCTGGGGCCAGCCTGGAGCCACTAGCCGGGCCGA 720
D 1192 CCCACCTGAGAGTAACTTTTGGATGGGGCTGGGGCCAGCCTGGAGCCACTAGCCGGGCCGA 1251
Qy 721 CTGACAGGCGATGAGCCCTGCGCTACCAACCCACATTTCTCAGATGACTGGCACTTCC 780
D 1252 CTGACAGGCGATGAGCCCTGCGCTACCAACCCACATTTCTCGGATGACTGGCACTTCC 1311
Qy 781 AGAGCCCTCCAGCAAGCAGCCCTTAGGATACAGCAGCCCTGTTTCCCTTCGCGGGGAAG 840
D 1312 AGAGCCCTCCAGCAAGCAGCCCTTAGGATACAGCAGCCCTGTTTCCCTTCGCGGGGAAG 1361
Qy 841 TCATAGGTTAGGAGCAGCTCACACTTCTTCAGGAGAGACACACAAACCATGACCCCTCA 900

Db	1362	-----	1361
QY	901	GCCTGGGACCCCAACTCCAGGCGCTCCAGCCCAACCTGCCAGCCAGCCCTGAAAT	960
Db	1362	-----GGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGAAAT	1401
QY	961	GCAAGTCTTGTACGAGTTTGAAGCTAGGAACCCACGGGAACCTGCTGTGTCAGGAGA	1020
Db	1402	GCAAGTCTTGTACGAGTTTGAAGCTAGGAACCCACGGGAACCTGCTGTGTCAGGAGA	1461
QY	1021	GAGCTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGCTGGTGAAGAATGAGCGGGAGC	1080
Db	1462	GAGCTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGCTGGTGAAGAATGAGCGGGAGC	1521
QY	1081	GAGCGCTACATTCACAGCAACATCTGGAGCCCTACAGCGGGAGCCCTGGGACCCA	1140
Db	1522	GAGCGCTACATTCACAGCAACATCTGGAGCCCTACAGCGGGAGCCCTGGGACCCA	1581
QY	1141	GGCCAGTCAACCTCTCGGCTTCCAATGCTTCAAGCTTAGCTCGAGGCTGAAGAGTCA	1200
Db	1582	GGCCAGTCAACCTCTCGGCTTCCAATGCTTCAAGCTTAGCTCGAGGCTGAAGAGTCA	1641
QY	1201	AGACTGGCTGACGAGAGAACTTCTCCACTGCGGCTGAGGACACTTGGTCCCTGAC	1260
Db	1642	AGACTGGCTGACGAGAGAACTTCTCCACTGCGGCTGAGGACACTTGGTCCCTGAC	1701
QY	1261	GGGAGCCAGTACTTCGCATAGACCTGGGAGCTACAGATGCTATGTCCACAGAGGC	1320
Db	1702	GGGAGCCAGTACTTCGCATAGACCTGGGAGCTACAGATGCTATGTCCACAGAGGC	1761
QY	1321	CCCAGAACTCTCTCCGGCTGGAGGCTGTGCAAGAGTGTGGGATAAGCCCTTAGGC	1380
Db	1762	CCCAGAACTCTCTCCGGCTGGAGGCTGTGCAAGAGTGTGGGATAAGCCCTTAGGC	1821
QY	1381	ACCAGTTAGACACTTCCAAAGACAGGCCCCCTGATGCAAGATGGCAGATCTGATAC	1440
Db	1822	ACCAGTTAGACACTTCCAAAGACAGGCCCCCTGATGCAAGATGGCAGATCTGATAC	1881
QY	1441	CATTAGAGCCCGAGAACTTCTCTCTGGATCCAGTGTGCAGCAACCCACACCCAG	1500
Db	1882	CATTAGAGCCCGAGAACTTCTCTCTGGATCCAGTGTGCAGCAACCCACACCCAG	1941
QY	1501	CTCACAGCAAAACAATGACAGGCCAGAGGCTGAAGCAAACTGTCTCTCTGGC	1560
Db	1942	CTCACAGCAAAACAATGACAGGCCAGAGGCTGAAGCAAACTGTCTCTCTGGC	2001
QY	1561	TGTGTTGGAGCCTCCCAAGTAAACCTATTTATTTTACCTCTTTCCCAACCTGGAGCA	1620
Db	2002	TGTGTTGGAGCCTCCCAAGTAAACCTATTTATTTTACCTCTTTCCCAACCTGGAGCA	2061
QY	1621	TTTATGCTAGGCTTGTCAAGAATCTGTTACAGTCCCTCTCTCTCAATAAAGCATCTT	1680
Db	2062	TTTATGCTAGGCTTGTCAAGAATCTGTTACAGTCCCTCTCTCTCAATAAAGCATCTT	2121
QY	1681	CAAGCTTGTAAAAAATAAANGA	1706
Db	2122	CAAGCTTGTAAAAAATAAANGA	2147
RESULT 4			
AY074930			
LOCUS			
DEFINITION			
Homo sapiens epidermal growth factor receptor pathway substrate 8			
related protein 3 (EPS8R3) mRNA, complete cds.			
ACCESSION			
AY074930			
VERSION			
AY074930.1 GI:18655334			
SOURCE			
Homo sapiens.			
ORGANISM			
Homo sapiens			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE			
1 (bases 1 to 1973)			

AUTHORS	Scita,G., Confalonieri,S., Offenhauser,N., Borgonovo,A.,									
TITLE	Tocchetti,A., Romano,P. and Di Fiore,P.P.									
JOURNAL	family									
REFERENCE	Unpublished									
AUTHORS	2 (bases 1 to 1973)									
TITLE	Tocchetti,A., Confalonieri,S., Offenhauser,N., Borgonovo,A.,									
JOURNAL	Direct Submission									
FEATURES	Submitted (23-JAN-2002) Experimental Oncology, European Institute									
source	of Oncology, Via Ripamonti 435, Milano 20141, Italy									
gene	1..1973									
CDS	/organism="Homo sapiens"									
	/db_xref="taxon:9606"									
	1..1973									
	/gene="EPS8R3"									
	112..1893									
	/gene="EPS8R3"									
	/note="EPS8 related protein 3"									
	/codon_start=1									
	/product="epidermal growth factor receptor pathway									
	substrate 8 related protein 3"									
	/protein_id="AAL76119.1"									
	/db_xref="GI:18655335"									
	/translation="MSRPSRAIYLHKEYSONLTSEPTLLQHRVHEHLMTCQGSORV									
	OGPEDALOKPEMDAQRWSDLILOVRDGLQLDLETKEELSYRLDSIQANMVA									
	LNTCSYNSILSITVOEPLPGTSTLLFOCEVGAERLTKSLQKALEELEQRPLGGL									
	QPSQWRGPMERPELMEQALIEPIPPPEQPHORTLEHSLPPSPRLPRTSAREP									
	SAFTLPPRRSSPEDDEVLNHLVDIELFMGKLEAKAKTSRKKFKKKNKQ									
	GGTQAQYIDCFQIKHSFNLLGLRLATWLKTSAPLTHILFKSLNFTLARGPEAGLA									
	AOVISPLLTAKAINLLQSCSPESNLWMLGPAWTTSRADWTGDEPLPYQPTFSDDW									
	OLPESSOAPILGYQDPVSLRSGHSLGSTSHFPOKETHNHDPPQDNPNSRSPSPKPAQ									
	PALMOLVLEFARNPRELTVVQGEKLEVLHDSKRWLVKNPAGSGSYIPSNILEPLO									
	PGTPTGQSPSRVPMRLSRSPPEVTWMLQAEENFSTATVTRLGSLTGSOLLIRPGE									
	IOMLCPOEAPRLSLREAVRMGLTSP"									
BASE COUNT	481 a	620 c	549 g	321 t	2 others					
ORIGIN										
Query Match	83.3%; Score 1424.4; DB 9; Length 1973;									
Best Local Similarity	99.3%; Pred. No. 0;									
Matches 1450; Conservative	0;	Mismatches	8;	Indels	2;	Gaps	2;			
QY	1	GGCAGAGCCACTGAAGACACAGCCTGCAGAGGCTCTGGAGGAAGAGCTGCAGCAAGACC	60							
Db	516	GGCAGAGCCACTGAAGACACAGCCTGCAGAGGCTCTGGAGGAAGAGCTGCAGCAAGACC	575							
QY	61	TGACCTTGGAGGCTTTCAGCCAGGCCAGGACAGATGGAGGGGCGCTGTATGGAAGGCC	120							
Db	576	TGACCTTGGAGGCTTTCAGCCAGGCCAGGACAGATGGAGGGGCGCTGTATGGAAGGCC	635							
QY	121	GCTCCCTATGGAGCAGCAGCTATCTGGAGCGGGGATCCCTCCAGACAGCCACCA	180							
Db	636	GCTCCCTATGGAGCAGCAGCTATCTGGAGCGGGGATCCCTCCAGACAGCCACCA	695							
QY	181	GAGGACCTTAGACACAGCCTCCACCATCCCAAGCCCTGCCACGCCACAGTGC	240							
Db	696	GAGGACCTTAGACACAGCCTCCACCATCCCAAGCCCTGCCACGCCACAGTGC	755							
QY	241	CCGAGAACCAAGTGCTTTTACTCTGCTCTCTCAAGCGGCTCTCTTCCCGAGGACC	300							
Db	756	CCGAGAACCAAGTGCTTTTACTCTGCTCTCTCAAGCGGCTCTCTTCCCGAGGACC	815							
QY	301	ACAGAGGGGACGAGGAGTCTGTAACCATGTCTTAAGGGACATTTGAGCTGTTCATGGGAA	360							
Db	816	ACAGAGGGGACGAGGAGTCTGTAACCATGTCTTAAGGGACATTTGAGCTGTTCATGGGAA	875							
QY	361	GCTGGAGAGGCCAGGCAAGACACAGCAGGAGGAAGAATAATTTGGGAAAAAACAGGA	420							
Db	876	GCTGGAGAGGCCAGGCAAGACACAGCAGGAGGAAGAATAATTTGGGAAAAAACAGGA	935							
QY	421	CCAGGAGGCTCTACCCAGGACAGTACATTTGACTCTCTCCAGAGATCAAGTACAGCTT	480							

Db 936 CCAGGGAGTCTACCCAGGCACAGTACATTGACTGCTTCCAGAAAGATCAAGCACAGCTT 995
Qy 481 CAACCTCTCTGGGAAGGTGGCCACCTGGCTGAAGGAGACAAGTCCCTCGAGTCGTACA 540
Db 996 CAACCTCTCTGGGAAGGTGGCCACCTGGCTGAAGGAGACAAGTCCCTCGAGTCGTACA 1055
Qy 541 CATCCTCTTCAAGTCCCTGAACATTCATCTCTGGCCAGGTGGCCCTGAGGCTAGCAGC 600
Db 1056 CATCCTCTTCAAGTCCCTGAACATTCATCTCTGGCCAGGTGGCCCTGAGGCTAGCAGC 1115
Qy 601 CCAAGTATCTACCCCTCTCTCACCCCTTAAGCTATCAACCTGCTACAGTCTCTCTAAG 660
Db 1116 CCAAGTATCTACCCCTCTCTCACCCCTTAAGCTATCAACCTGCTACAGTCTCTCTAAG 1175
Qy 661 CCCACCTGAGAGTAACCTTTTGATGGGTTGGGCCAGCCTGGACCCTAGCCGGCCGA 720
Db 1176 CCCACCTGAGAGTAACCTTTTGATGGGTTGGGCCAGCCTGGACCCTAGCCGGCCGA 1235
Qy 721 CTGGACAGGGGATGAGCCCTTGCCTTACCAACCCACATTTCTCAGATGACTGGCAACTTCC 780
Db 1236 CTGGACAGGGGATGAGCCCTTGCCTTACCAACCCACATTTCTCAGATGACTGGCAACTTCC 1295
Qy 781 AGAGCCCTCCAGCCCAAGCACCTTAGGATACCAAGACCTGCTTTCCCTTCGGCCGGGAAG 840
Db 1296 AGAGCCCTCCAGCCCAAGCACCTTAGGATACCAAGACCTGCTTTCCCTTCGGCCGGGAAG 1355
Qy 841 TCATAGGTTAGGGAGCACCTCACACTTTCCCTCAGGAGAAGACACAAACCATGACCTCA 900
Db 1356 TCATAGGTTAGGGAGCACCTCACACTTTCCCTCAGGAGAAGACACAAACCATGACCTCA 1415
Qy 901 GCCTGGGSAACCCCACTCCAGGCCCTCCAGCCCAACCTTCCAGCCAGCCCTGAAAT 960
Db 1416 GCCTGGGSAACCCCACTCCAGGCCCTCCAGCCCAACCTTCCAGCCAGCCCTGAAAT 1475
Qy 961 GCAAGCTTTGACAGTTTGAAGCTTAGAAGACCCAGGGAACTGACTGTGTCAGGAGA 1020
Db 1476 GCAAGCTTTGACAGTTTGAAGCTTAGAAGACCCAGGGAACTGACTGTGTCAGGAGA 1535
Qy 1021 GAAGCTGGAGTCTTGGACCAACAGCAAGCGTGTGCTGGTGAAGATGAGCGGGACG 1080
Db 1536 GAAGCTGGAGTCTTGGACCAACAGCAAGCGTGTGCTGGTGAAGATGAGCGGGACG 1595
Qy 1081 GAGCGGCTTACATTCGAAGCAACATCTCTGGAGCCCTACAGCCGGGACCCCTGGGACCCA 1140
Db 1596 GAGCGGCTTACATTCGAAGCAACATCTCTGGAGCCCTACAGCCGGGACCCCTGGGACCCA 1655
Qy 1141 GGGCCAGTCAACCTCTCGGGTTCCAATGCTTCGACTTAGCTCGAGGCTGAAGAGGTAC 1200
Db 1656 GGGCCAGTCAACCTCTCGGGTTCCAATGCTTCGACTTAGCTCGAGGCTGAAGAGGTAC 1715
Qy 1201 AGACTGGCTGAGGAGAGAACTTCTCCACTGCCACGGTGAAGACACTTGGGTCCCTGAC 1260
Db 1716 AGACTGGCTGAGGAGAGAACTTCTCCACTGCCACGGTGAAGACACTTGGGTCCCTGAC 1775
Qy 1261 GGGGAGCCAGCTACTTCGCATTAAGACCTTGGGAGCTACAGATGCTATGTCACAGGAGGC 1320
Db 1776 GGGGAGCCAGCTACTTCGCATTAAGACCTTGGGAGCTACAGATGCTATGTCACAGGAGGC 1835
Qy 1321 CCCAGGAATCTGTCGGGCTGGAGGCTGTCAGAAAGATGCTGGGGATAGCCCTTAGGC 1380
Db 1836 CCCAGGAATCTGTCGGGCTGGAGGCTGTCAGAAAGATGCTGGGGATAGCCCTTAGGC 1894
Qy 1381 ACCAGCTTAGACACTTCCAAGAACACAGCCCGCTGATGCAAGATGGCAGATCTGATACC 1440
Db 1895 ACCAGCTTAGACACTTCCAAGAACACAGCCCGCTGATGCAAGATGGCAGATTTGATACC 1933
Qy 1441 CATTAGAGCCCGGAGAAATTC 1460
Db 1954 CATTANAGCCCGGAGAAATTC 1973

RESULT 5
AX322735

LOCUS AX322735 2159 bp DNA linear PAT 07-JAN-2002
DEFINITION Sequence 10 from Patent WO0192528.
ACCESSION AX322735
VERSION AX322735.1 GI:18093725
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1
AUTHORS Macina, R. A.; Chen, S. Y.; Pluta, J.; Sun, Y. and Recipon, H.
TITLE Method of diagnosing, monitoring, staging, imaging and treating
colon cancer
JOURNAL Patent: WO 0192528-A 10 06-DEC-2001;
diadexus, Inc. (US)
FEATURES
source Location/Qualifiers
1..2159
/organism="Homo sapiens"
/db_xref="taxon:9606"
BASE COUNT 528 a 673 c 581 g 377 t
ORIGIN
Query Match 78.0%; Score 1334.6; DB 6; Length 2159;
Best Local Similarity 92.0%; Pred. No. 0;
Matches 1580; Conservative 0; Mismatches 29; Indels 108; Gaps 12;
Qy 1 GGCAGAGCGACTGAAGACCAGCCCTGCAGAGAGGCTCTGGAGAGAGAGCTGGAGCAAGACC 60
Db 532 GGCAGAGCGACTGAAGACCAGCCCTGCAGAGAGGCTCTGGAGAGAGAGCTGGAGCAAGACC 591
Qy 61 TCGACTTTGGAGGCTTCAGCCAGGCCAGGACAGATGAGAGGGGCTGCTATATGGAAGGCC 120
Db 592 TCGACTTTGGAGGCTTCAGCCAGGCCAGGACAGATGAGAGGGGCTGCTATATGGAAGGCC 651
Qy 121 GCTCCCTATGGAGCAGCAGCCCTATCTGGAGCCCGGGATCCCTCCAGACAGCCCAACCA 180
Db 652 GCTCCCTATGGAGCAGCAGCCCTATCTGGAGCCCGGGATCCCTCCAGACAGCCCAACCA 711
Qy 181 GAGACCCCTAGAGCACAGCCCTCCACCATCCCAAGGCCCTGCCAGGCCACACAGTGC 240
Db 712 GAGACCCCTAGAGCACAGCCCTCCACCATCCCAAGGCCCTGCCAGGCCACACAGTGC 771
Qy 241 CCGAGAACCAGTGCCCTTTACTCTGCTCTCTCCAAAGGGGTCCTCTTCCGCCAGGAGACC 300
Db 772 CCGAGAACCAGTGCCCTTTACTCTGCTCTCTCCAAAGGGGTCCTCTTCCGCCAGGAGACC 831
Qy 301 AGAGAGGAGCAGGAGTGTCTGAACCATGTCTTAAGGACATTCAGCTGTTTCATGGGAAA 360
Db 832 AGAGAGGAGCAGGAGTGTCTGAACCATGTCTTAAGGACATTCAGCTGTTTCATGGGAAA 891
Qy 361 GCTGGAGAAGGCCAGCAAGACAGCAGCAAGAGAAATTTGGGAAAAA---AACAA 417
Db 892 GCTGGAGAAGGCCAGCAAGACAGCAGCAAGAGAAATTTGGGAAAAAAGAGAGAACAA 951
Qy 418 GGACCCAGGAGGTCTACCCAGGCACAGTACA-TTGACTGCTTCCAGAGATCAAGTACA 476
Db 952 GGACCCAGGAGGTCTACCCAGGCACAGTACAGTTGACTGCTTCCAGAGATCAAGCACA 1011
Qy 477 GCTTCAACCTCTGGAGGCTGGCCACCTGGCTGAAGGAGACAGTGCCTTGAAGTCG 536
Db 1012 GCTTCAACCTCTGGAGGCTGGCCACCTGGCTGAAGGAGACAGTGCCTTGAAGTCG 1071
Qy 537 TACACATCTCTTCAAGTCCCTGAACCTTCATCTTGGCAGGTGCCCTGAGCTGGCCCTAG 596
Db 1072 TACACATCTCTTCAAGTCCCTGAACCTTCATCTTGGCAGGTGCCCTGAGCTGGCCCTAG 1131
Qy 597 CAGCCCAAGTGATCTACCCCTCTCAACCTTAAGCTATCAACCTGCTACAGTCTCTGTC 656
Db 1132 CAGCCCAAGTGATCTACCCCTCTCAACCTTAAGCTATCAACCTGCTACAGTCTCTGTC 1191
Qy 657 TAAGCCCACTGAGAGTAACCTTTGGATGGGTTGGGCCAGCCTGGACCACTAGCCGGG 716
Db 1192 TAAGCTCACCTGAGAGTAACCTTTGGATGGGTTGGGCCAGCCTGGACCACTAGCCGGG 1251

QY	717	CCGACTGGACAGCGATGAGCCCTGCGCTTACCAACCCACATTTCTCAGATGACTGGCAAC	776
Db	1252	CCGACTGGACAGCGATGAGCCCTGCGCTTACCAACCCACATTTCTCAGATGACTGGCAAC	1311
QY	777	TTCCAGAGCCCTCCAGCCAGCACCCTTAGGATACACAGGACCCCTGTTCCCTTCGCGCGG	836
Db	1312	TTCCAGAGCCCTCCAGCCAGCACCCTTAGGATACACAGGACCCCTGTTCCCTTCGCGCGG	1366
QY	837	GAAGTCATAGTTAGGAGCACCTCACACTTCTCAGGAGAGACACACCAACCATGACC	896
Db	1367	-----	1366
QY	897	CTCAGCCTGGGACCCCAACTCCAGGCCCTTCCAGCCCAAACTGCCCAACCCAG-CCCTG	955
Db	1367	-----GGCCTCCAGTCCCAAACTGCCCAACCCAGTCCCTG	1402
QY	956	AAATGCAAGTCTGTACGAGTTTGAAGCTAGGAA--CCACGGGAACTGACTGTGGTCC	1013
Db	1403	AAATGCAAGTCTGTACGAGTTTGAAGCTAGGAA--CCACGGGAACTGACTGTGGTCC	1462
QY	1014	AGGAGAGAGCTGGAGGTTCTGACACAGCAAGCGGTGGTGGTCAAGAAATGAGG	1073
Db	1463	AGGAGAGAGCTGGAGGTTCTGACACAGCAAGCGGTGGTGGTCAAGAAATGAGG	1522
QY	1074	CGGACGGAGCGGCTACATTTCCAAAGCAACATCTGGAGCCCTTACAGCGGGGACCCCTG	1133
Db	1523	CGGACGGAGCGGCTACATTTCCAAAGCAACATCTGGAGCCCTTACAGCGGGGACCCCTG	1582
QY	1134	GGACCCAGGCGCAGTCA-CCCTTCGGGTTTCAATGCTTGGACTTACGTCGAGGCCCTGAA	1192
Db	1583	GGACCCAGGCGCAGTCA-CCCTTCGGGTTTCAATGCTTGGACTTACGTCGAGGCCCTGAA	1642
QY	1193	GAGTCAACAGCTGGCTGAGGAGAGACTTCTCAGTCCACGTCGAGGACACTTGGG	1252
Db	1643	GAGTCAACAGCTGGCTGAGGAGAGACTTCTCAGTCCACGTCGAGGACACTTGGG	1702
QY	1253	TCCTGACGCGG--AGCCAGCTACTTCGCA-TAAGACCTGGGGAGGTAC--AGATGCTAT	1307
Db	1703	TCCTGACGCGGAGCCAGCTACTTCGCAATTAAGACCTGGGGAGCTACCGAGTCTAT	1762
QY	1308	GTCACAGGAGGCGCCAC---GAATCTGTCCGGCTGAGGCTGTGAGAGGATGC-TG	1363
Db	1763	GTCACAGGAGGCGCCAC---GAATCTGTCCGGCTGAGGCTGTGAGAGGATGC-TG	1822
QY	1364	GGATAGCCCTTAGGCACACGCTTAGACACCTTCCAGAACACGCGCCGCTGATGCAAG	1423
Db	1823	GGATAGCCCTTAGGCACACGCTTAGACACCTTCCAGAACACGCGCCGCTGATGCAAG	1882
QY	1424	ATGGCAGATCTGATCCCATTTAGAGCCCGGAGAAATTCCTTCTTGGATCCCGTTTGCAG	1483
Db	1883	ATGGCAGATCTGATCCCATTTAGAGCCCGGAGAAATTCCTTCTTGGATCCCGTTTGCAG	1942
QY	1484	CAAAACCCACAC--CCAGCTCACACAGCAAAACAAATGACAGGCGCCAGAGCTGAAGC	1541
Db	1943	CAAAACCCACACCTCCAGCGCTCACACAGCAAAACAAATGACAGGCGCCAGAGCTGAAGC	2002
QY	1542	AAACAGTGTCCCTTCTGGCTGTCTTGGAGCTCCCGAGTCCCAAGAACCACTATTATTTACCT	1601
Db	2003	AAACAGTGTCCCTTCTGGCTGTCTTGGAGCTCCCGAGTCCCAAGAACCACTATTATTTACCT	2062
QY	1602	CTTTCCCAACCTGGAGCATTTATGCTAGGCTTGTCAAGAAATCTGTTAGTCCCTCTCC	1661
Db	2063	CTTTCCCAACCTGGAGCATTTATGCTAGGCTTGTCAAGAAATCTGTTAGTCCCTCTCC	2122
QY	1662	TTCTCAATAAAGCATCTTCAAGCTTGTAAAAA 1698	
Db	2123	TTCTCAATAAAGCATCTTCAAGCTTGTAAAAA 2159	
RESULT 6			
BC014734			
LOCUS			

DEFINITION	Mus musculus, Similar to hypothetical protein FLJ21522, clone
ACCESSION	MGC:25893 IMAGE:4218079, mRNA, complete cds.
VERSION	BC014734
KEYWORDS	BC014734.1 GI:15928516
SOURCE	MGC.
ORGANISM	house mouse.
REFERENCE	Mus musculus.
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 2275)
TITLE	Strausberg, R.
JOURNAL	Direct Submission
REMARK	Submitted (01-OCT-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
COMMENT	NIH-MGC Project URL: http://mgc.nci.nih.gov Contact: MGC help desk Email: cgapbs-re@mail.nih.gov Tissue Procurement: Jeffrey E. Green, M.D. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center Center code: BCM-HGSC Web site: http://www.hgsc.bcm.tmc.edu/cdna/ Contact: amg@bcm.tmc.edu Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.
FEATURES	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Series: IRAK Plate: 31 Row: b Column: 17 This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein. Location/Qualifiers 1..2275 /organism="Mus musculus" /db_xref="taxon:10090" /map="FVB/N" /clone="MGC:25893 IMAGE:4218079" /tissue_type="Colon, normal. 5 month old male mouse." /clone_lib="NCI_CGAP_Co24" /lab_host="DH10B" /note="Vector: pCMV-SPORT6" 150..1952 /codon_start=1 /product="Similar to hypothetical protein FLJ21522" /protein_id="AAH14734.1" /db_xref="GI:15928517" /translation="MSRPSRAIYLHKEYSOSMASEPTLLOHVRHLMTCIKLGTORV REPFDALQKLEMDAGRWVSDLFQVRDGLWLLDIEETKEELDSYRLDNLKAIIDVA LNTCSYLSITVQESGLPSTLLFQCEYAGQLRTSLQALEELEERPFVGH HPSQDRWKGPPLERPLPQQAPLEQRFSPHPPPEQPHNMTSPRSISPSRSILTHY PSARENGFTLPPRPAPSPEDPERDEVLNHLRDLIELFAGKLKVEQARNSHKTK LGRKKKSKNGITQAEYIDCFQIKLSNLLGKLALRMOETSAPFVGLIFQTLKFL SQCPCAGLPKAVISPELLTPKADLLQSCLSPPEDTLWKSGLGTSWTSWADWTGSEPPP YQPTFYDQWQIPQPSMPPITNQDSISLGRSMRSLHPPRDEPNYHNPEYEDSNLPL SSPSGRAALKMVLVEFARNABLTVAQGEILVLDQSKRWLVKKEAGLTGYTPS NLTEPLPAGAPRGHQPFRAPMLRLSSKPEVTAWLQAEFSTVTVRTLGLSMGSQL LHMRPEQLMCLCPQEPAPRQARLDVRRMLGTH"
CDS	BASE COUNT 620 a 565 g 439 t ORIGIN Query Match 53.7%; Score 918; DB 10; Length 2275; Best Local Similarity 74.4%; Pred. No. 1.4e-237; Matches 1294; Conservative 0; Mismatches 396; Indels 49; Gaps 9;
QY	1 GGCAGAGCAGCTGAAGACGAGCTGCAGAGGCTCTTGGAGGAGAGCTGGAGCAAGACC 60
Db	554 GGCAGAGCAGCTGGAGCAAGCCTGCAGAGGCCCTGGAAGAGAGCTAGAGCAAGACC 613


```
/gene="Eps8R3"
/notes="Eps8 Related protein 3"
/codon_start=1
/product="epidermal growth factor receptor pathway
substrate 8 related protein 3"
/protein_id="AAU76121.1"
/db_xref="GI:18655339"
/translation="MSRPSRAIYLHRKEYSOSMAPEPLLQHRVHLMTCILGTQRV
RPKDALQIQMDAQGRVMSQDLFQVRDGLHLDIETKEELDSYRLDNIKAIDVA
LNTCSVNSILSVTVDSGLPGISTLLFCQVEYGAELRTSLOKALEEELERPRGVH
HPSODRWKGPPLERPLIQOAPLEQFSEPHRFPEQPHNMTSRSISPSRSLTHY
PARENGTLLPPPRAPSPDEPDEVLNHRLDIEFAGKLEQVQARNHKKTK
LGRKKKNGITQABYIDCFQIKLSFNLLKALRMOETSAPEFVGLFIQFLKPI
SCQPEAGLPKIVISPLLPKALDILQSLSPEDTLWKSLSGTSWTTSWADWTGSEPP
YQTFYDQGIQIPRSMPITNQDSISLGRSMRSLHPRDEPNHNPEYEDSNLPL
SSPSQRAAKMOVLVEFARNAOELTVAGQLEVLVDQSKRWLVKNEAGLTGYIPS
NILEPLPACAPRGRHPSFRAPMLRLSSKPEVTAWLQAFNESTVTVRTLGLSIMG
LHMRGELQMLCPQEPAPRIQARLDVARMGLMTH"
BASE COUNT      648 a   659 c   574 g   448 t
ORIGIN
Query Match      53.4%; Score 913.2; DB 10; Length 2329;
Best Local Similarity 74.2%; Pred. No. 2.8e-236;
Matches 1291; Conservative 0; Mismatches 399; Indels 49; Gaps 9;
Qy 1 GGCAGAGCGACTGAAGACCGAGCTGCAGAAAGGCTCTGGAGAAAGAGCTGGAGCAAGACC 60
Db 582 GGCAGAGCAGCTGAGGACAAAGCGCTGCAGAAAGGCCCTGGAAGAAGAGAGCTAGAAGAAAGACC 641
Qy 61 TCGACTTTGGAGGCTTCAGCCAGCCAGGACACAGATGGAGGGGGCGCTGCTATGAAAGGCC 120
Db 642 TCGATTTGGAGTCCATCATCCAGCCAGGACAGATGGAAGGGGCGCTCTCTGTGAAAGGCC 701
Qy 121 GCTCCCTATGGAGCAGCGCTATCTGGAGCGGGGATC-----CC 162
Db 702 ACTCCCTATACAGCGCCACCCCTCTGGAGCAGAGGTTCTCTCCAGAGCATAGGTTTCC 761
Qy 163 TCCAGAACAGCCCAACAGAGACCCCTTAGACACAGCTCCCAACCATCCCAAGGCCCT 222
Db 762 TCCAGAACAGCCACATAACATGACCTCAGAACAGCAGCATCTGCCCATCTCTCAAGTCTCT 821
Qy 223 GCCAGCCACACAGTCCCGCAGAACCAAGTGCGTTTACTCTG---CCTCCTCAAGGCG 279
Db 822 GACACACTACCCCAAGTCCCGCAGAACCAATGGCTTCACTCTGCCTCTCTCCCGAGGCG 881
Qy 280 GTCTCTCTCCCGAGGACCCAGAGAGGGACGAGGAGTGCTGAACCATCTCTCTAAGGGA 339
Db 882 TGCTCCCTCTCCCGAGGACCCAGAGAGGGATGAGGAGTGCTGAACCATCTCTCTAAGAGA 941
Qy 340 CATTTGAGCTGTTTCATGGGAAAGCTGGAGAAGGCCCGCCAGCAAGACAGCAGGAAGAGAA 399
Db 942 CATTTGAGCTGTTTCTGGAAGCTGAAGGAGTCCAGGCAAGAACAAGTCTATAAGAGAC 1001
Qy 400 A-----TTTGGGAAAAAACAAGGACCAAGGAGGTTCTCACCCAGGCAAGTACATTTGA 453
Db 1002 AAAACTTTGGGAGAAAAAAGAAAGCTAATAAATGGGATACACAAAGCAGAGTACATTTGA 1061
Qy 454 CTGCTTCCAGAGATCAAGTACAGCTTCAACCTCTCTGGGAGGCTGCCACCTGGCTGAA 513
Db 1062 CTGCTTCCAGAGATCAAGTACAGCTTCAACCTCTCTGGGAGGCTGCCACCTGGGATTTGA 1121
Qy 514 GGAGACAAGTGCCCTGAGCTTACACATCTCTCTCAAGTCCCTGAACCTTCACTCTGGC 573
Db 1122 GGAACAAGCGCCCTGAGTTCGTGGCCCTCATCTTCAAAACCTGAAATTTATCTATC 1181
Qy 574 CAGTGCCCTGAGGCTGGCTAGAGCCCAAGTATCTACCCCTCTCTACCCCTCAAGC 633
Db 1182 TCAGTGCCCTGAGGCTGGCTTTCAGCAAGGATTTCCCTCCCTCTCTACCCCTCAAGC 1241
Qy 634 TATCAACCTGCTACAGTCTCTTAAGCCACCTGAGAGTAACTTTGGATGGGTTGGG 693
Db 1242 AATAGACCTGCTGAGTCTGTAAAGCCACCCGAGGACACACTCTGGAAGTCACTTAG 1301
```

RESULT 8
AX407643/c
LOCUS
DEFINITION

Sequence 290 from Patent W00229103.

401 bp

DNA

linear

PAT 14-JUN-2002

```
Qy 694 CCAGCCTGGACACACTAGCGGGCGGCGACTGGACAGCGGATGAGCCCTGCCCTACCAACC 753
Db 1302 CACATCTGGACACCACTGGCTGGCTGACTGGACAGCGAGTGAACCGCACCTTACCAACC 1361
Qy 754 CACATCTCAGATGACTGGCAACTTCCAGAGCCCTCCAGCAAGCACCCCTTAGATACCA 813
Db 1362 CACATCTCAGATGACTGGCGAGATTCACAGCCCGCTCCATGATGCCCATACCAACC 1421
Qy 814 GGACCTGTTTCCCTTCTGGCGGGGAAGTCACTAGTTAGGGAGCACCTTCACACTTTCTCTCA 873
Db 1422 GGATTCGATTTCCCTCC-----GAGTTCTAGATGAGGAGCGCTAGATTTCCCTCG 1475
Qy 874 GGAGAAGACACACAACTATGACCTCAGCCTGGGGACCCCAACTCCAGGCCCTCCAGCCC 933
Db 1476 GGATGAGCCATACCAACCAATAACTGAGGAGCTCAAACTCCCACTCTCCAGCCC 1535
Qy 934 CAAACCTGCCACCGCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCAAGGAATGC 993
Db 1536 CAGCCCTGGCAGAGCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCAAGGAATGC 1595
Qy 994 ACGGGAACCTGACTGTGTCCAGGGAGAGAGCTGGAGTTCTGGACACAGCAAGCGGTG 1053
Db 1596 CCAGGAACCTGACTGTGCACAGGGGAGATTTCTGGAGTTCTGATCAGAGCAAGCGATG 1655
Qy 1054 GTGGCTGGTGAAGATGAGGGGAGCGGAGCGGCTACATTTCCAGCAACATCTCTGGAGCC 1113
Db 1656 GTGGCTGGTGAAGAAAGCAAGCAGAGCTGACTGGTTTACATTTCCCAAGCAATATCTCTGGAGCC 1715
Qy 1114 CCTACAGCCGGGACCCCTGGGACCCAGGCGCAGTCACTCTCGGTTTCCAATGCTTCG 1173
Db 1716 CTTGCCAGCTGGAGCCCGGAGGCCACCAACCACTCTTTAGGGCCCAATGCTTCG 1775
Qy 1174 ACTTAGCTCGAGCGCTGAAGAGGTCACAGAGCTGGCTGCAGGAGAGAACTTCTCCACTGC 1233
Db 1776 CTTGAGCTCAAGGCTGAGGAGGTCACAGCTGGCTGCAGGAGAGAACTTCTCCACTGT 1835
Qy 1234 CACGGTGAGGACACTTGGGTCTCTGACGGGGAGCCAGCTACTTTCGATAAGACTTGGGGA 1293
Db 1836 CACTGTAAGCACTCTGGGTCTCTTAATGGGAAGCAGCTACTTTCACATGAGACTGGGGA 1895
Qy 1294 GCTACAGATCTATGTCCAGAGGAGCCCAAGATCTCTCCCGGCTGGAGCTGTCTAG 1353
Db 1896 GCTGCAGATCTGTGTCCAGAGGAGGCTCCACGGATCCAGGGCCCGGCTGGAGCTGTCTAG 1955
Qy 1354 AAGGATCTGGGATTAAGCCCTTAGCCACAGCTTAGACACCTCCAAAGAACACAGG-CCCC 1412
Db 1956 AAGGATCTGGGATTAAGCTTAGGAGCAGCTCGATACCTTACGATACGAGCAAGGAGCTC 2015
Qy 1413 GCTGATGCAAGATGGCAGATCTGATACCCA--TTAGAGCCCGGAGAAATTCCTTCTTGGAT 1471
Db 2016 TCATGTGCAAGATGGAGTATTTGGTACCCAGTTAGAGCCCTCGAAGTCTCTCTT----- 2069
Qy 1472 CCAGTTTGCAGAAACCCCAACCCAGCTCAGACAGGCAAAACAANTGACAGGCCCCAG 1531
Db 2070 -TCAGTTAGTGGCAAGCCACATATCTTGGATCAACAAGCAAGAAAGATGAACAGTCCAG 2128
Qy 1532 AGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGAGCTTCCCAAGTAAACCACTATT 1591
Db 2129 AACTGAGGCAAAATGGTACCTCATTTGGTTG---GAGACCTTCAACCGTTA----CTATT 2181
Qy 1592 TATTTTACCTCTTTTCCCAAACTGGAGCATTTATGCTTAGGCTTTGTAAGAATCTCTGTCA 1651
Db 2182 TATTTTACCCACACCCCTAAGTTGGCATACTTATTTCCAGTCTTACCTCTGATCTCTTTA 2241
Qy 1652 GTCCCTCTCTCTCAATAAAGCATCTTCAAGCTTGTAAATAAATAAATAAANGATAAA 1710
Db 2242 GTCCCAATCTGCAATAAAGCAATCGTTTAACTTGTCTATGAAAAAATAAATAAATAA 2300
```


[illegible]


```

/note="LRR7 repeat: matches 1. .450 of consensus"
15622. .15992
/note="MER4-internal repeat: matches 1502. .1866 of
consensus"
15995. .16388
/note="MER4-internal repeat: matches 836. .1268 of
consensus"
16466. .16505
/note="MER4-internal repeat: matches 687. .726 of
consensus"
16510. .16805
/note="AluSc repeat: matches 1. .296 of consensus"
16815. .16888
/note="MER4-internal repeat: matches 751. .824 of
consensus"
17226. .17859
/note="MER4-internal repeat: matches 1. .692 of consensus"
17863. .18321
/note="MER4A2 repeat: matches 1. .504 of consensus"
18590. .18710
/note="L2 repeat: matches 2601. .2725 of consensus"
18829. .19149
/note="MLT1I repeat: matches 16. .363 of consensus"
19332. .19625
/note="AluX repeat: matches 1. .294 of consensus"
19659. .20116
/note="LRR2 repeat: matches 2. .449 of consensus"
20173. .20260
/note="MIR repeat: matches 2. .91 of consensus"
20328. .20469
/note="L2 repeat: matches 2131. .2282 of consensus"
21358. .21860
/note="MLT1E repeat: matches 1. .539 of consensus"
21876. .22075
/note="MER61D repeat: matches 342. .743 of consensus"
22851. .22889
/note="LJM3e repeat: matches 515. .556 of consensus"
23193. .23489
/note="AluJo repeat: matches 1. .296 of consensus"
23563. .23648
/note="LIME repeat: matches 5683. .5757 of consensus"
23649. .23955
/note="AluSc repeat: matches 1. .307 of consensus"
23956. .24093
/note="LIME repeat: matches 5497. .5683 of consensus"
24386. .24441
/note="MER41A repeat: matches 400. .452 of consensus"
26554. .26853
/note="Charlie4a repeat: matches 227. .508 of consensus"
26854. .27170
/note="AluJb repeat: matches 1. .312 of consensus"
27171. .27348
/note="Charlie4a repeat: matches 2. .227 of consensus"
27745. .28699
/note="CpG Island"
29257. .29603
/evidence=not_experimental
/note="L2 repeat: matches 2424. .2750 of consensus"
29892. .30198
/note="L2 repeat: matches 1322. .1655 of consensus"
30614. .30739
/note="MIR repeat: matches 137. .261 of consensus"
30741. .30788
/note="24 copies 2 mer tt 75% conserved"
30802. .36956
/note="LipA4 repeat: matches 6. .6146 of consensus"
36958. .37084
/note="MIR repeat: matches 23. .151 of consensus"
41990. .42041
/note="L2 repeat: matches 2652. .2703 of consensus"
42995. .43450
/note="LMB4 repeat: matches 5731. .6181 of consensus"
43538. .43843

```

```

repeat_region /note="AluJb repeat: matches 1. .290 of consensus"
44248. .44545
repeat_region /note="AluSq repeat: matches 7. .306 of consensus"
47160. .47197
repeat_region /note="19 copies 2 mer gt 100% conserved"
50429. .50483
repeat_region /note="L2 repeat: matches 2443. .2497 of consensus"
50653. .50853
repeat_region /note="MIR repeat: matches 18. .233 of consensus"
53782. .53851
repeat_region /note="L2 repeat: matches 2685. .2749 of consensus"
53853. .54465
repeat_region /note="L2 repeat: matches 1749. .2413 of consensus"
54631. .55246
repeat_region /note="LRR5 repeat: matches 1. .623 of consensus"
55326. .55629
repeat_region /note="LRR5 repeat: matches 675. .969 of consensus"
55714. .55843
repeat_region /note="L2 repeat: matches 2387. .2516 of consensus"
55897. .55938
repeat_region /note="21 copies 2 mer tg 81% conserved"
56457. .56528
repeat_region /note="L2 repeat: matches 2387. .2459 of consensus"
56728. .57841
repeat_region /note="L2 repeat: matches 1389. .2609 of consensus"
59949. .60010
repeat_region /note="31 copies 2 mer gt 80% conserved"
60563. .60721
repeat_region /note="MER5B repeat: matches 18. .177 of consensus"
61673. .61733
repeat_region /note="MIR repeat: matches 104. .164 of consensus"
61971. .61994
repeat_region /note="MIR repeat: matches 117. .140 of consensus"
62581. .62627
repeat_region /note="MIR repeat: matches 200. .240 of consensus"
62628. .62901

Query Match 19.0%; Score 324.4; DB 9; Length 133451;
Best Local Similarity 99.7%; Pred. No. 1.6e-76;
Matches 325; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1366 GATAAGCCCTTAGCACCAGCTTAGACACCTCCAAGAACACGAGCCCGCTGATGCAAGAT 1425
Db 38418 GATAAGCCCTTAGCACCAGCTTAGACACCTCCAAGAACACGAGCCCGCTGATGCAAGAT 38359

QY 1426 GGCAGATCTCATACCCATTAGAGCCCGAGAAATTCCTCTTCTGGATCCCAGTTGCAGCA 1485
Db 38358 GGCAGATCTCATACCCATTAGAGCCCGAGAAATTCCTCTTCTGGATCCCAGTTGCAGCA 38299

QY 1486 AACCCACACCCAGCTCACACAGCAAAACAAATGGACGCCAGAGGCTGAAGCAAAAC 1545
Db 38298 AACCCACACCCAGCTCACACAGCAAAACAAATGGACGCCAGAGGCTGAAGCAAAAC 38239

QY 1546 AGTGTCCCTTCTGGCTGTGTGGAGCCTCCCGAGTAACCACTATTTATTTTACCTCTTT 1605
Db 38238 AGTGTCCCTTCTGGCTGTGTGGAGCCTCCCGAGTAACCACTATTTATTTTACCTCTTT 38179

QY 1606 CCCAACCTGGAGCATTTATGCCTTAGCTTGTCAAGAAATCTGTTCAGTCCCTCTCTCTCT 1665
Db 38178 CCCAACCTGGAGCATTTATGCCTTAGCTTGTCAAGAAATCTGTTCAGTCCCTCTCTCTCT 38119

QY 1666 CAATAAAGCATCTTCAAGCTTTGTAA 1691
Db 38118 CAATAAAGCATCTTCAAGCTTTGTCA 38093

RESULT 10
LOCUS AC092860 162480 bp DNA linear HTG 19-AUG-2002
DEFINITION Pan troglodytes clone rpd43-125j14, WORKING DRAFT SEQUENCE, 5
ordered pieces.
ACCESSION AC092860
VERSION AC092860.16 GI:22297381

```

KEYWORDS
SOURCE
ORGANISM

HTG; HTGS_PHASE2; HTGS_DRAFT.
chimpanzee.
Pan troglodytes
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.

REFERENCE
AUTHORS
TITLE
JOURNAL
AUTHORS
TITLE
JOURNAL

1 (bases 1 to 162480)
Pan troglodytes BAC Clone rp43-125j14
Unpublished
2 (bases 1 to 162480)
Jiang,X., Song,L. and Roe,B.A.
Direct Submission
Submitted (31-JUL-2001) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

3 (bases 1 to 162480)
Jiang,X., Song,L., Eichler,E. and Roe,B.A.
Direct Submission
Submitted (19-AUG-2002) Department of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

COMMENT

On Aug 19, 2002 this sequence version replaced gi:22218546.

Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR

* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* provided by the submittor.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 2375: contig of 2375 bp in length
* 2376 2475: gap of unknown length
* 2476 16668: contig of 14193 bp in length
* 16669 16768: gap of unknown length
* 16769 35334: contig of 18766 bp in length
* 35335 35635: gap of unknown length
* 35635 92812: contig of 57178 bp in length
* 92813 92913: gap of unknown length
* 92913 162480: contig of 69568 bp in length.
* Location/Qualifiers
1..162480
/organism="Pan troglodytes"
/db_xref="taxon:9598"
/clone_lib="rp43-125j14"
/clone_lib="RPCI - 43 Male Chimpanzee BAC Library"
BASE COUNT 42683 a 39383 c 38740 g 41229 t 445 others
ORIGIN

Query Match 18.9%; Score 322.8; DB 2; Length 162480;
Best Local Similarity 99.4%; Pred. No. 4.5e-76;
Matches 324; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1366 GATAGCCCTTAGGCACGCTTAGACACCTCCCAAGAACACGAGCCGCCCTGATGCAAGAT 1425
DB 44411 GATAGCCCTTAGGCACGCTTAGACACCTCCCAAGAACACGAGCCGCCCTGATGCAAGAT 44470

QY 1426 GGCAGATCTGATACCCATTAGAGCCGCCGAGAAATTCCTCTTGGATCCAGCTTGCAGCA 1485
DB 44471 GGCAGATCTGATACCCATTAGAGCCGCCGAGAAATTCCTCTTGGATCCAGCTTGCAGCA 44530

QY 1486 AACCCACACCCGCTCACAGCAGCAAAACAAATGGACAGGCCCGCAGAGCGCTCAAGCAAAAC 1545
DB 44531 AACCCACACCCGCTCACAGCAGCAAAACAAATGGACAGGCCCGCAGAGCGCTCAAGCAAAAC 44590

QY 1546 AGTGCCCTCTGGCTGTGGTGGAGCCCTCCCGAGTAACACCTATTATTATTACCTCTTT 1605
DB 44591 AGTGCCCTCTGGCTGTGGTGGAGCCCTCCCGAGTAACACCTATTATTATTACCTCTTT 44650

QY 1606 CCCAAACCTGGAGCATTATATCCCTAGGCTTGTCAAGAAATCTGTTCAGTCCCTCCTCTTCT 1665
DB 44651 CCCAAACCTGGAGCATTATATCCCTAGGCTTGTCAAGAAATCTGTTCAGTCCCTCCTCTTCT 44710
QY 1666 CAATAAAAGCATCTTCAAGCTTGTA 1691
DB 44711 CAATAAAAGCATCTTCAAGCTTGTA 44736

RESULT 11
G25672

LOCUS human STS EST47775, sequence tagged site. 224 bp DNA linear STS 02-JUN-1996
ACCESSION G25672
VERSION G25672.1 GI:1347904
KEYWORDS STS; STS sequence; primer; sequence tagged site.
SOURCE Homo sapiens STSS derived from sequences in dbEST and the Unigene collection.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 224)

REFERENCE
AUTHORS
TITLE

Hudson,T.
Whitehead Institute/MIT Center for Genome Research; Physically
Mapped STSS

JOURNAL
COMMENT

Unpublished (1995)
Contact: Thomas Hudson
Whitehead Institute/MIT Center for Genome Research
Whitehead Institute for Biomedical Research
9 Cambridge Center, Cambridge MA 02142 USA
Tel: 617 252 1900
Fax: 617 252 1902
Email: thudson@genome.wi.mit.edu

Primer A: GTTTCAGCAAAACCCAC
Primer B: TCCAGGTTGGGAAAGAGGT
STS size: 142
PCR Profile:
Presoak:
Denaturation:
Annealing: 56 degrees C
Polymerization:
PCR Cycles: 35
Thermal Cycler:
Protocol:
Template: 10 ng
Primer: each 5 pM
dNTPs: each 4 nM
Taq Polymerase: 0.025 units/ul
Total Vol: 20 ul

Buffer:
MgCl2: 1.5 mM
KCl: 50 mM
Tris-HCL: 10 mM
pH: 9.3

Derived from dbEST (genbank accession D25742).
Location/Qualifiers
1..224
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="406.4 cR from top of Chr1 linkage group"
STS
primer_bind 8..25
BASE COUNT 61 a 67 c 37 g 55 t 4 others
ORIGIN

Query Match 12.8%; Score 218.4; DB 11; Length 224;
Best Local Similarity 97.8%; Pred. No. 5.5e-48;

```
Matches 219; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1469 GATCCAGTTTGCAGCAAAACCCACACCCAGCTCACAGCAAAACCAATGGACAGGC 1528
Db 1 GATCCAGTTTGCAGCAAAACCCACACCCAGCTCACAGCAAAACCAATGGACAGGC 60
QY 1529 CAGAGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGAGCCTCCCGAGTAACCACT 1588
Db 61 CAGAGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGAGCCTCCCGAGTAACCACT 120
QY 1589 ATTATTTTACCTCTTCCAAACCTGGAGCATTTATGCTAGGCTGTCAAGAAATCTGT 1648
Db 121 ATTATTTTACCTCTTCCAAACCTGGAGCATTTATGCTAGGCTGTCAAGAAATCTGT 180
QY 1649 TCAGTCCCTCTCCTTCTCAATAAAGCATCTTCAAGCTTGTAAA 1692
Db 181 TCAGTCCCTCTCCTTCTCAATAAAGCATCTTCAAGCTTGTAAA 224

RESULT 12
G41866
LOCUS SHG3-32779 Human Homo sapiens STS genomic, sequence tagged site.
DEFINITION G41866
ACCESSION G41866.1 GI:3668199
VERSION
KEYWORDS
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 224)
AUTHORS Myers,R.M.
TITLE Human STSs (1998)
JOURNAL Unpublished (1998)
COMMENT
Contact: Richard M. Myers
Stanford Human Genome Center (SHGC)
Stanford University School of Medicine
Department of Genetics, M-344, Stanford, CA 94305, USA
Tel: 4157259687
Fax: 4157259689
Email: myers@hgc.stanford.edu
Primer A: GTTTCGACGAAACCCAC
Primer B: TCCAGGTTTGGGAAAGAGT
STS size: 142
PCR Profile:
Initial incubation: 95 degrees C for 10 minutes
Denaturation: 94 degrees C for 30 seconds
Annealing: 60 degrees C for 30 seconds
Polymerization: 72 degrees C for 23 seconds
PCR Cycles: 30
Thermal Cycler: Perkin Elmer 9700
Protocol:
Template: 25 ng
Primer: each 1 uM
dNTPs: each 200 uM
Amplifaq Gold Polymerase: 0.07 units/ul
Total Vol: 5 ul
Buffer: MgCl2: 2.5 mM
KCl: 50 mM
Tris-HCl: 10 mM
pH: 8.3
Prepared with primer pairs provided by Sandoz, derived from D25742
-- Washington University/Merck EST sequence.
FEATURES
source
1..224
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="11"
```

```
STS /clone_lib="Human"
primer_bind 8..149
primer_bind 8..25
primer_bind complement(130..149)
BASE COUNT 61 a 67 c 37 g 55 t 4 others
ORIGIN
Query Match 12.8%; Score 218.4; DB 11; Length 224;
Best Local Similarity 97.8%; Pred No. 5.5e-48;
Matches 219; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 1469 GATCCAGTTTGCAGCAAAACCCACACCCAGCTCACAGCAAAACCAATGGACAGGC 1528
Db 1 GATCCAGTTTGCAGCAAAACCCACACCCAGCTCACAGCAAAACCAATGGACAGGC 60
QY 1529 CAGAGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGAGCCTCCCGAGTAACCACT 1588
Db 61 CAGAGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGAGCCTCCCGAGTAACCACT 120
QY 1589 ATTATTTTACCTCTTCCAAACCTGGAGCATTTATGCTAGGCTGTCAAGAAATCTGT 1648
Db 121 ATTATTTTACCTCTTCCAAACCTGGAGCATTTATGCTAGGCTGTCAAGAAATCTGT 180
QY 1649 TCAGTCCCTCTCCTTCTCAATAAAGCATCTTCAAGCTTGTAAA 1692
Db 181 TCAGTCCCTCTCCTTCTCAATAAAGCATCTTCAAGCTTGTAAA 224

RESULT 13
AR044148
LOCUS Sequence 40 from patent US 5817479.
DEFINITION AR044148
ACCESSION AR044148
VERSION AR044148.1 GI:5965613
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 197)
AUTHORS Au-Young,J., Bandman,O., Hawkins,P.R. and Wilde,C.G.
TITLE Human Kinase homologs
JOURNAL Patent: US 5817479-A 40 06-OCT-1998;
FEATURES Location/Qualifiers
source 1..197
/organism="unknown"
BASE COUNT 50 a 52 c 63 g 32 t
ORIGIN
Query Match 9.1%; Score 155.6; DB 6; Length 197;
Best Local Similarity 96.0%; Pred. No. 5.8e-31;
Matches 192; Conservative 0; Mismatches 4; Indels 4; Gaps 3;
QY 932 CCAAAACCTGCCAGCCAGCCCTGAAAATGCAAGTCTTTGTACGAGTTTGAAGCTAGGAAC 991
Db 1 CCAAAACCTGCCAGCCAGCCCTGAAAATGCAAGTCTTTGTACGAGTTTGAAGCTAGGAAC 60
QY 992 CCACGGGAAGTACTGCTGTGTCAGGGAGAGAGCTGGAGTCTTGTGACGAGTTGACGACGACG 1051
Db 61 CCACGGGAAGTACTGCTGTGTCAGGGAGAGAGCTGGAGTCTTGTGACGAGTTTGAAGCTAGGAAC 119
QY 1052 TGGTGGCTGTGAAGAAATGAGCGGAGCGGCTACATTCGAAGCAACATCTCTGGAG 1111
Db 120 TGGTGGCTGTGAAGAAAT-AGCGGGAGCGGAGCGGCTACATTCGAAGCAACATC--TGGG 176
QY 1112 CCCTTACAGCGGGGACCC 1131
Db 177 CCCTTACAGCGGGGACCC 196

RESULT 14
AC018461/c
LOCUS AC018461
DEFINITION Mus musculus chromosome 3 clone mgs1-192g21 strain 129/SvJ,
```

```

complete sequence.
AC018461
VERSION AC018461.30 GI:21427705
KEYWORDS HTG:
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Ray,L., Qi,S., Loh,P., Perng,T., Carson,M., Pearson,W. and Roe,B.A.
TITLE Mus musculus BAC Clone mgs1-192q21
JOURNAL Unpublished
REFERENCE
AUTHORS Ray,L., Qi,S., Loh,P., Pearson,B. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (11-DEC-1999) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Ray,L., Qi,S., Loh,P., Pern,T., Carson,M., Pearson,W. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (07-JUN-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Ray,L., Qi,S., Loh,P., Pern,T., Carson,M., Pearson,W. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (15-JUN-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Jun 15, 2002 this sequence version replaced gi:21328313.
-----
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
FEATURES
source
Location/Qualifiers
1. .113285
/organism="Mus musculus"
/strain="129/SvJ"
/db_xref="taxon:10090"
/chromosome="3"
/clone="mgs1-192q21"
/clone_lib="Genome Systems mouse ES cell BAC libraries I
and II"
BASE COUNT 28787 a 26952 c 27495 g 30051 t
ORIGIN
Query Match 6.9%; Score 117.6; DB 10; Length 113285;
Best Local Similarity 75.0%; Pred. No. 1.9e-20;
Matches 147; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 837 GAAGTCATAGTTAGGGAGCACCCTCACACTTCTCTCAGGAGAGACACACACCATGACC 896
DB 43033 GAGGTTCTAGGATGAGGAGCAGCCTACATTTCCCTCGGGATGAGGCATACACCATACC 42974
QY 897 CTCACCTCGGGAGCCCAACTCCAGGCCTCCAGCCCCCAAACTCCCGAGCCGACCTGA 956
DB 42973 CTGAATACGAGGACTCAAACTCCCACTCCAGCCCCCAGCCCTGGCAGAGCAGCCCTGA 42914
QY 957 AAATGCAAGTCTTTACGAGTTTGAAGCTAGGAACCCCGGAACTGACTGTGGTCCAGG 1016
DB 42913 AAATGCAAGTCTGTACGAGTTTGAAGCAAGGAATGCCAGGAAGTACTGTGGCACAGG 42854
QY 1017 GAGAGAAGCTGGAGGT 1032
DB 42853 GGGAGATTCTGGAGGT 42838
RESULT 15
AC079042 198459 bp DNA linear HTG 20-AUG-2002
LOCUS AC079042
DEFINITION Mus musculus chromosome 3 clone rp23-313f7 strain C57BL/6J, WORKING

```

```

ACCESSION AC079042
VERSION AC079042.27 GI:22325017
KEYWORDS HTG: HTGS_PHASE2; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS Sloan,D., Hua,A., Hong,X., Pearson,W. and Roe,B.A.
TITLE Mus musculus BAC Clone rp23-313f7
JOURNAL Unpublished
REFERENCE
AUTHORS Sloan,D., Hua,A., Hong,X., Pearson,W. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (17-AUG-2000) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS Sloan,D., Hua,A., Hong,X., Pearson,W. and Roe,B.A.
TITLE Direct Submission
JOURNAL Submitted (20-AUG-2002) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Aug 20, 2002 this sequence version replaced gi:22218516.
-----
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 2 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 57584: contig of 57584 bp in length
* 57585 57684: gap of unknown length
* 57685 198459: contig of 140775 bp in length.
FEATURES
source
Location/Qualifiers
1. .198459
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/chromosome="3"
/clone="rp23-313f7"
/clone_lib="RPCI - 23 Female (C57BL/6J) Mouse BAC Library"
BASE COUNT 55177 a 47384 c 46269 g 49494 t 135 others
ORIGIN
Query Match 6.9%; Score 117.6; DB 2; Length 198459;
Best Local Similarity 75.0%; Pred. No. 2e-20;
Matches 147; Conservative 0; Mismatches 49; Indels 0; Gaps 0;
QY 837 GAAGTCATAGTTAGGGAGCACCCTCACACTTCTCTCAGGAGAGACACACACCATGACC 896
DB 128837 GAGGTTCTAGGATGAGGAGCAGCCTACATTTCCCTCGGGATGAGGCATACACCATACC 128896
QY 897 CTCAGCCTCGGGAGCCCCAACTCCAGGCCTCCAGCCCCCAAACTGCCAGCCGACCTGA 956
DB 128897 CTGAATACGAGGACTCAAACTCCCACTCCAGCCCCCAGCCCTGGCAGAGCAGCCCTGA 128956
QY 957 AAATGCAAGTCTTTACGAGTTTGAAGCTAGGAACCCCGGAACTGACTGTGGTCCAGG 1016
DB 128957 AAATGCAAGTCTTTACGAGTTTGAAGCAAGGAATGCCAGGAAGTACTGTGGCACAGG 129016
QY 1017 GAGAGAAGCTGGAGGT 1032
DB 129017 GGGAGATTCTGGAGGT 129032

```

Search completed: February 25, 2003, 08:44:52
Job time : 3704 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 05:25:18 ; Search time 261 Seconds
(without alignments)
14754.472 Million cell updates/sec

Title: US-09-762-021a-1

Perfect score: 1710

Sequence: 1 ggcagagcgactgaagacca.....aaaaaaaaaataagataaa 1710

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1709	99.9	1710	21	AAZ95379 Human colon specif
2	1696.6	99.2	2220	22	AAF32641 Human cDNA encodin
3	1334.6	78.0	2159	24	ABA91311 Human colon specif
4	832.8	48.7	893	23	ABK43674 DNA encoding novel
5	830.8	48.6	872	22	AAH33515 Human colon cancer
6	829.6	48.5	872	22	AAH33515 Human colon cancer
7	829.6	48.5	872	22	AAH33515 Human colon cancer
8	673.8	39.4	1265	22	AAH33515 Human colon cancer
9	671.8	39.3	1282	23	ABK43673 DNA encoding novel

c	10	389	22.7	401	24	ABN93792	Gene #290 used to
	11	388.8	22.7	1388	22	AAK88048	Human digestive sy
	12	335.4	19.6	572	22	AAH27487	cDNA encoding nove
	13	335.4	19.6	572	23	ABK43956	DNA encoding novel
	14	331	19.4	786	20	AAZ16535	Human gene expres
c	15	324	18.9	2643	22	AAH27685	DNA encoding novel
	16	297.8	17.4	338	22	AAH34654	Human colon cancer
	17	218.4	12.8	224	16	AAH22498	Human gene signatu
	18	180.2	10.5	481	22	AAH36539	Human colon cancer
	19	157.4	9.2	6784	22	AAH89660	Human digestive sy
	20	155.6	9.1	197	19	AAV33556	Clone 237002 cDNA
	21	119.2	7.0	216	21	AAH98311	Human colon cancer
	22	95.8	5.6	2457	22	AAH99536	Human protein enco
	23	93.2	5.5	1635	22	AAH24282	Human Ovi107 ovar
	24	86.4	5.1	2535	22	AAH94459	Human full-length
	25	68.8	4.0	595	21	AAH98568	Human colon cancer
	26	62.6	3.7	2922	23	ABL18887	Drosophila melanog
	27	62.6	3.7	3031	23	ABL07407	Drosophila melanog
	28	62.6	3.7	6013	23	ABL18686	Drosophila melanog
	29	62.6	3.7	8511	23	ABL07406	Human spliced tran
	30	60	3.5	60	24	ABN41901	DNA encoding nove
	31	56.4	3.3	1638	23	ABK43695	cDNA encoding nove
	32	56.4	3.3	1639	22	AAH27064	Human ovarian and
	33	56.4	3.3	1661	22	AAH07672	Human reproductive
	34	56.4	3.3	1661	22	AAH02553	cDNA encoding nove
	35	56.4	3.3	1661	22	AAH27477	DNA encoding novel
	36	56.4	3.3	1661	23	ABK43975	Human ovarian anti
	37	56.4	3.3	1663	24	ABQ54695	Human cDNA sequenc
	38	56.4	3.3	1947	22	AAH14778	Lung small cell ca
	39	55	3.2	514	24	AAH61502	DNA encoding novel
	40	53.8	3.1	1655	23	AAH33902	Mouse eps8. Mus m
	41	49.8	2.9	3245	14	AAQ40730	Gene encoding a su
	42	49.8	2.9	10732	21	AAH10594	Human Nck-2 cDNA.
	43	47.4	2.8	1457	22	AAH2642	Human KDR signal t
	44	47.4	2.8	2267	20	AAH84665	Drosophila melanog
c	45	46.6	2.7	9232	23	ABL20448	

ALIGNMENTS

RESULT 1
AAZ95379

ID AAZ95379 standard; DNA; 1710 BP.

AC AAZ95379;

XX

DT 01-JUN-2000 (first entry)

DE Human colon specific gene (CSG) SEQ ID NO.1.

DE Human colon specific gene (CSG) SEQ ID NO.1.

KW Human; colon specific gene; CSG; diagnosis; colon cancer; detection;
KW cytosolic; ds.

OS Homo sapiens.

PN WO200007632-A1.

XX

PD 17-FEB-2000.

PF 20-JUL-1999; 99WO-US16357.

XX

PR 04-AUG-1998; 98US-0095231.

XX

PA (DIAD-) DIADEXUS LLC.

XX

PI Sun Y, Recipon H, Macina RA;

XX

DR WPI; 2000-205579/18.

XX

PT Novel methods for diagnosing, monitoring, staging, imaging and treating

XX

PT colon cancer by measuring the level of colon specific gene markers -

XX

PS Claim 6; Page 37-38; 42pp; English.

XX The present invention describes a method for diagnosing the presence of
CC colon cancer in a patient. The method comprises measuring levels of
CC colon specific gene markers (CSG) in cells, tissues or bodily fluids,
CC and comparing the measured levels of CSG with levels of CSG from a
CC normal human control, where an increase in measured CSG levels in the
CC patient versus control is associated with the presence of colon cancer.
CC AA295379 to AA295381 represent human CSG sequences. The method is used
CC to detect, monitor, stage or give a prognosis for colon cancer.
CC Antibodies against CSGs are used for detection or image localisation of
CC the CSGs. The antibodies can be conjugated to cytotoxic agent or drug and
CC used to treat colon cancer. The method of the invention is more accurate
CC than prior art clinical methods for staging colon cancer, because it
CC measures colon specific markers, and, unlike pathological staging
CC methods, do not depend on an invasive procedure.

XX Sequence 1710 BP; 437 A; 529 C; 444 G; 299 T; 1 other;

Query Match 99.9%; Score 1709; DB 21; Length 1710;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1710; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GCACGAGCGACTGAAGACAGCCTGCAGAGGCTCTGGAGGAAGAGCTGGAGCAAGACC	60
DB	1	GCACGAGCGACTGAAGACAGCCTGCAGAGGCTCTGGAGGAAGAGCTGGAGCAAGACC	60
QY	61	TCGACTTGGAGGCCCTTCACGACAGGCCAGGACAGATGGAGGGGCCCTGCTATGGAAGGCC	120
DB	61	TCGACTTGGAGGCCCTTCACGACAGGCCAGGACAGATGGAGGGGCCCTGCTATGGAAGGCC	120
QY	121	GCTCCTATGGAGCGACGCTATCTGGAGCCGGGGATCCCTCCAGAACAGCCCCACCA	180
DB	121	GCTCCTATGGAGCGACGCTATCTGGAGCCGGGGATCCCTCCAGAACAGCCCCACCA	180
QY	181	GAGGACCTTAGACAGCGCTCCACCATCCCAAGGGCCCTGCCACGCCACACCACTGC	240
DB	181	GAGGACCTTAGACAGCGCTCCACCATCCCAAGGGCCCTGCCACGCCACACCACTGC	240
QY	241	CCGAGAACCAAGTGCTTACTCTGCTCTCCAAAGGGGCTCTCTCCCGGAGGACCC	300
DB	241	CCGAGAACCAAGTGCTTACTCTGCTCTCCAAAGGGGCTCTCTCCCGGAGGACCC	300
QY	301	AGAGAGGGACGAGGAAGTGTCTGAACCATGTCTTAAGGGACATTTAGCTGTTCATGGGAA	360
DB	301	AGAGAGGGACGAGGAAGTGTCTGAACCATGTCTTAAGGGACATTTAGCTGTTCATGGGAA	360
QY	361	GCTGGAGAGGCCCGGCAAGAACACAGGAGGAGAGAAATTTGGGAAAAAACAAGGA	420
DB	361	GCTGGAGAGGCCCGGCAAGAACACAGGAGGAGAGAAATTTGGGAAAAAACAAGGA	420
QY	421	CCAGGAGGCTCTCACCCAGGACAGTACATTCAGCTGCTTCCAGAGATCAAGTACAGCTT	480
DB	421	CCAGGAGGCTCTCACCCAGGACAGTACATTCAGCTGCTTCCAGAGATCAAGTACAGCTT	480
QY	481	CAACCTCTCGGAAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGGCCCTGAGCTCGTACA	540
DB	481	CAACCTCTCGGAAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGGCCCTGAGCTCGTACA	540
QY	541	CATCTCTTCAAGTCCCTGAACCTTCACTTGGCCAGGTGCCCTGAGGCTGGCCTAGCAGC	600
DB	541	CATCTCTTCAAGTCCCTGAACCTTCACTTGGCCAGGTGCCCTGAGGCTGGCCTAGCAGC	600
QY	601	CCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGCTACAGTCTCTGTCTAAG	660
DB	601	CCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGCTACAGTCTCTGTCTAAG	660
QY	661	CCCACCTGAGAGTAACCTTTGGATGGGTTGGGCCCGCAGCCTGGACCACTAGCCGGCGGA	720
DB	661	CCCACCTGAGAGTAACCTTTGGATGGGTTGGGCCCGCAGCCTGGACCACTAGCCGGCGGA	720
QY	721	CTGGACAGCGGATGAGCCCTCTGCCCTACCAACCCACATTCCTCAGATGACTGGCAACTTCC	780

DB	721	CTGGACAGCGGATGAGCCCTCTGCCCTACCAACCCCAATTCCTCAGATGACTGGCAACTTCC	780
QY	781	AGAGCCCTCCAGCCCAAGCACCTTAGGATACAGGACCCCTGTTTCCCTTCCGCGGGGAAG	840
DB	781	AGAGCCCTCCAGCCCAAGCACCTTAGGATACAGGACCCCTGTTTCCCTTCCGCGGGGAAG	840
QY	841	TCATAGGTTAGGAGCACCTCACACTTCTCAGGAGAAGACACACAACCATGACCCCTCA	900
DB	841	TCATAGGTTAGGAGCACCTCACACTTCTCAGGAGAAGACACACAACCATGACCCCTCA	900
QY	901	GCCTGGGACCCCAACTCCAGGCCCTCCAGCCCAAACTGCCACGAGCCCTCAAAAT	960
DB	901	GCCTGGGACCCCAACTCCAGGCCCTCCAGCCCAAACTGCCACGAGCCCTCAAAAT	960
QY	961	GCAAGCTTTGTACAGAGTTTGAAGCTTAGGAACCCACGGGAACCTGACTGTGGTCCAGGGAGA	1020
DB	961	GCAAGCTTTGTACAGAGTTTGAAGCTTAGGAACCCACGGGAACCTGACTGTGGTCCAGGGAGA	1020
QY	1021	GAACTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGTGGTGAAGAATGAGCGGGGACG	1080
DB	1021	GAACTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGTGGTGAAGAATGAGCGGGGACG	1080
QY	1081	GAGCGGCTACATTCCAAGCAACATCCTGGAGCCCTACAGCGGGGACCCCTGGGACCCA	1140
DB	1081	GAGCGGCTACATTCCAAGCAACATCCTGGAGCCCTACAGCGGGGACCCCTGGGACCCA	1140
QY	1141	GGCCAGTCAACCTCTCGGGTTCCAAATGCTTCGACTTACGCTCGAGGCTTGAAGAGGTAC	1200
DB	1141	GGCCAGTCAACCTCTCGGGTTCCAAATGCTTCGACTTACGCTCGAGGCTTGAAGAGGTAC	1200
QY	1201	AGACTGGCTGAGGAGAGAACTTCTCCACTGCCAGCTGAGGACACTTTGGTCCCTGAC	1260
DB	1201	AGACTGGCTGAGGAGAGAACTTCTCCACTGCCAGCTGAGGACACTTTGGTCCCTGAC	1260
QY	1261	GGGAGCCAGCTACTTCGCATAAGACCTGGGAGGTACAGATGCTATGTCCACAGGAGGC	1320
DB	1261	GGGAGCCAGCTACTTCGCATAAGACCTGGGAGGTACAGATGCTATGTCCACAGGAGGC	1320
QY	1321	CCCAGAACTCTGTCCTCCGCTGGAGGCTGTGAGAGGATGCTGGGGATAAGCCCTTAGGC	1380
DB	1321	CCCAGAACTCTGTCCTCCGCTGGAGGCTGTGAGAGGATGCTGGGGATAAGCCCTTAGGC	1380
QY	1381	ACCAGCTTAGACACCTCCAAAGAACCCAGCCCGCTGTGATGCAAGATGGCAGATCTGATACC	1440
DB	1381	ACCAGCTTAGACACCTCCAAAGAACCCAGCCCGCTGTGATGCAAGATGGCAGATCTGATACC	1440
QY	1441	CATTAGAGCCCGGAGAACTTCTCTTCTGATCCCAAGTTTGGCAGCAACCCCAACCCCAAG	1500
DB	1441	CATTAGAGCCCGGAGAACTTCTCTTCTGATCCCAAGTTTGGCAGCAACCCCAACCCCAAG	1500
QY	1501	CTCACACAGCAAAAACAATGGACAGCCCGCAGAGGCTGAAGGCTGAAGCAACAGTGTCTGCGC	1560
DB	1501	CTCACACAGCAAAAACAATGGACAGCCCGCAGAGGCTGAAGGCTGAAGCAACAGTGTCTGCGC	1560
QY	1561	TGTGTTGAGGCTTCCCAAGTAAACCACTATTTATTTTACCTCTTTCCCAAACTCGAGCA	1620
DB	1561	TGTGTTGAGGCTTCCCAAGTAAACCACTATTTATTTTACCTCTTTCCCAAACTCGAGCA	1620
QY	1621	TTTATGCTTACGCTTGTCAAGAACTGTTCACTCCCTCTCTCTCAATTAAGCATCTT	1680
DB	1621	TTTATGCTTACGCTTGTCAAGAACTGTTCACTCCCTCTCTCTCAATTAAGCATCTT	1680
QY	1681	CAAGCTTGTAAAAAATAAANGATATAA 1710	
DB	1681	CAAGCTTGTAAAAAATAAANGATATAA 1710	

RESULT 2
AAF32641
ID AAF32641 standard; cDNA; 2220 BP.
XX
AC AAF32641;
XX

22-MAR-2001 (first entry)
Human cDNA encoding intracellular signalling molecule INTRA4.
Human; intracellular signalling molecule; INTRA; immunosuppressive; cytostatic; neuroprotective; nontropic; antiarteriosclerotic; cancer; antiinflammatory; anti-HIV; neuroleptic; antibacterial; antifungal; antiviral; antiparasitic; antihelminthic; antiparkinsonian; AIDS; cell proliferative disorder; arteriosclerosis autoimmunity; epilepsy; inflammatory disorder; Addison's disease; gastrointestinal disorder; neurological disorder; Parkinson's disease; Creutzfeldt-Jakob disease; mental disorder; schizophrenia; anxiety; ss.
Homo sapiens.
WO200077040-A2.
21-DEC-2000.
16-JUN-2000; 2000WO-US16636.
16-JUN-1999; 99US-0139566.
17-AUG-1999; 99US-0149640.
09-NOV-1999; 99US-0164417.
(INCY-) INCYTE GENOMICS INC.
Yue H. Tang YT, Hillman JL, Lal P, Bandman O, Baughn MR;
Azimzai Y, Yang J, Reddy R, Lu DM;
WPT; 2001-025334/03.
P-PSDB; AAB64372.
New human intracellular signaling molecules, useful for the diagnosis, prevention and treatment of cell proliferative, autoimmune, inflammatory, neurological, gastrointestinal, reproductive and developmental disorders -
Claim 5; Page 162; 192pp; English.
Sequences AAF32638 - AAF32689 represent cDNA encoding human intracellular signalling molecules INTRA1 - INTRA52, represented in AAB64369 - AAB64420. Modulators of the intracellular signalling molecules of the invention exhibit immunosuppressive; cytostatic; neuroprotective; neuroleptic; antiarteriosclerotic; antiinflammatory; anti-HIV; antihelminthic; antibacterial; antifungal; antiviral; antiparasitic; agonists and antagonists are useful for the treatment of a condition associated with decreased or increased expression of functional INTRA. Disorders associated with abnormal INTRA expression or activity include cell proliferative disorders e.g. arteriosclerosis and cancers; autoimmune or inflammatory disorders e.g. Addison's disease and acquired immunodeficiency syndrome (AIDS); viral, bacterial, fungal, parasitic, protozoal and helminthic infections; gastrointestinal disorders e.g. dysphagia and irritable bowel syndrome; neurological disorders e.g. epilepsy and Parkinson's disease; prion diseases e.g. Creutzfeldt-Jakob disease and mental disorders e.g. anxiety, schizophrenia and Tourette's disorder. Antibodies immuno specific for the INTRA proteins may also be useful in the diagnosis of the above disorders.
Sequence 2220 BP; 549 A; 696 C; 593 G; 382 T; 0 other;
Query Match 99.2%; Score 1696.6; DB 22; Length 2220;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1699; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
Qy 1 GGCAGAGCGACTGAAGACCGCTGCAGAGGCTCTGGAGGAAGAGCTGGAGCAAGACG 60
Db 518 GGCAGAGCGACTGAAGACCGCTGCAGAGGCTCTGGAGGAAGAGCTGGAGCAAGACG 577
Qy 61 TCGACTTGGAGGCTTCAGCCAGGCGGACAGATGAGGAGGGGCTGCTATGGAAAGGCC 120
Db 578 TCGACTTGGAGGCTTCAGCCAGGCGGACAGATGAGGAGGGGCTGCTATGGAAAGGCC 637

Qy 121 GCTCCCTATGGAGCAGCAGCGCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCCACCA 180
Db 638 GCTCCCTATGGAGCAGCAGCGCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCCACCA 697
Qy 181 GAGGACCTTAGAGCAGCAGCGCTCCACCATCCCAAGGCCCTGCCAGCCACACAGTGC 240
Db 698 GAGGACCTTAGAGCAGCAGCGCTCCACCATCCCAAGGCCCTGCCAGCCACACAGTGC 757
Qy 241 CCGAGAACCAGTGCCTTTACTCTGCTCCCTCCCAAGGGGCTCTTCCCCCGAGGAGCC 300
Db 758 CCGAGAACCAGTGCCTTTACTCTGCTCCCTCCCAAGGGGCTCTTCCCCCGAGGAGCC 817
Qy 301 AGAGGGAGCAGGAGAGTGTCTGAACCATGTCTTAAGGAGACATTTGAGTGTTCATGGGAAA 360
Db 818 AGAGGGAGCAGGAGAGTGTCTGAACCATGTCTTAAGGAGACATTTGAGTGTTCATGGGAAA 877
Qy 361 GCTGGAGAGGCCCGAGCAAGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 420
Db 878 GCTGGAGAGGCCCGAGCAAGACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 937
Qy 421 CCAGGGAGGTCTCACCCAGGACAGTACATTTGAGTGTCTCCAGAGAGATCAAGTACAGCTT 480
Db 938 CCAGGGAGGTCTCACCCAGGACAGTACATTTGAGTGTCTCCAGAGAGATCAAGTACAGCTT 997
Qy 481 CAACCTCTGGGAGAGCTGGCCACTGGCTGAAGGAGACAAAGTCCCTGAGCTCGTACA 540
Db 998 CAACCTCTGGGAGAGCTGGCCACTGGCTGAAGGAGACAAAGTCCCTGAGCTCGTACA 1057
Qy 541 CATCTCTTCAAGTCCCTGAACCTTCATCTGGCCAGGTGGCCCTGAGGCTGGCCCTAGCAGC 600
Db 1058 CATCTCTTCAAGTCCCTGAACCTTCATCTGGCCAGGTGGCCCTGAGGCTGGCCCTAGCAGC 1117
Qy 601 CCAAGTGTCTCACCCCTCTCCACCCCTAAAGCTATCAACCTGTACAGTCTGTCTTAAG 660
Db 1118 CCAAGTGTCTCACCCCTCTCCACCCCTAAAGCTATCAACCTGTACAGTCTGTCTTAAG 1177
Qy 661 CCCACCTGAGAGTAACCTTTGGATGGGTTGGGCCAGCTGGGACCACTAGCCGGGCCGA 720
Db 1178 CCCACCTGAGAGTAACCTTTGGATGGGTTGGGCCAGCTGGGACCACTAGCCGGGCCGA 1237
Qy 721 CTGACAGGCGATGAGGCGCTTCCCTACCAACCCACATTTCTCAGATGACTGGCAACTTCC 780
Db 1238 CTGACAGGCGATGAGGCGCTTCCCTACCAACCCACATTTCTCAGATGACTGGCAACTTCC 1297
Qy 781 AGAGCCCTCCAGCCCAAGCAGCCCTTAGATACCAGGACCTGTCTTCCCTTGGCGGGGAAG 840
Db 1298 AGAGCCCTCCAGCCCAAGCAGCCCTTAGATACCAGGACCTGTCTTCCCTTGGCGGGGAAG 1357
Qy 841 TCATAGCTTAGGAGCAGCCTCACACTTTCTCTCAGGAGAGACACACACACATGACCCCTCA 900
Db 1358 TCATAGCTTAGGAGCAGCCTCACACTTTCTCTCAGGAGAGACACACACACATGACCCCTCA 1417
Qy 901 GCCTGGGAGCCCAACTTCCAGGCGCTCCAGCCCAACCTGCCAGCCAGCCCTGAAAT 960
Db 1418 GCCTGGGAGCCCAACTTCCAGGCGCTCCAGCCCAACCTGCCAGCCAGCCCTGAAAT 1477
Qy 961 GCAAGTCTTACAGGTTTGAAGCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1020
Db 1478 GCAAGTCTTACAGGTTTGAAGCTAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1537
Qy 1021 GAAGCTGGAGGTTCTGGAGCAGCAGAGCGGTGGTGGCTGGTGAAGATGAGGCGGGGAGC 1080
Db 1538 GAAGCTGGAGGTTCTGGAGCAGCAGAGCGGTGGTGGCTGGTGAAGATGAGGCGGGGAGC 1597
Qy 1081 GAGCGGTACATTCACAGCAACATCTCTGGAGCCCTTACAGCCGGGGAGCCCTTGGAGCCCA 1140
Db 1598 GAGCGGTACATTCACAGCAACATCTCTGGAGCCCTTACAGCCGGGGAGCCCTTGGAGCCCA 1657
Qy 1141 GGCCAGCTACCTCTCGGGTTCCAAATGCTTGCATTTAGCTCGAGGCGCTGAGAGGCTCAC 1200
Db 1658 GGCCAGCTACCTCTCGGGTTCCAAATGCTTGCATTTAGCTCGAGGCGCTGAGAGGCTCAC 1717

1201 AGACTGGCTGCAGGAGAACTTCTCCACTCCCGGTGAGGACACTTGGTCCCTGAC 1260
1718 AGACTGGCTGCAGGAGAGAACTTCTCCACTCCCGGTGAGGACACTTGGTCCCTGAC 1777
1261 GGGGAGCCAGCTACTTCCGATAGACCTTGGGAGCTACAGATGCTATGTCCACAGAGGC 1320
1778 GGGGAGCCAGCTACTTCCGATAGACCTTGGGAGCTACAGATGCTATGTCCACAGAGGC 1837
1321 CCACGAAATCCCTGCTCCCGCTGAGGCTGTCTCAGAGGATGCTGGGATAGCCCTTAGGC 1380
1838 CCACGAAATCCCTGCTCCCGCTGAGGCTGTCTCAGAGGATGCTGGGATAGCCCTTAGGC 1897
1381 ACCAGCTTAGACACCTTCCAAAGAACAGGCCCGCTGATGCAAGATGGCAGATCTGATACC 1440
1898 ACCAGCTTAGACACCTTCCAAAGAACAGGCCCGCTGATGCAAGATGGCAGATCTGATACC 1957
1441 CATTTAGAGCCCGAGAAATTCCTTCTTGTGATCCAGTTTTCAGCAAAACCCACACCCAG 1500
1958 CATTTAGAGCCCGAGAAATTCCTTCTTGTGATCCAGTTTTCAGCAAAACCCACACCCAG 2017
1501 CTCACACAGCAAAACAAATGGACAGGCCAGAGGCTGAAGCAAAACAGTGTCCCTTCTGCG 1560
2018 CTCACACAGCAAAACAAATGGACAGGCCAGAGGCTGAAGCAAAACAGTGTCCCTTCTGCG 2077
1561 TGTGTTGGAGCCTCCCGCTGAGTACCACTATTTATTTTACCTCTTTCCCAAACTGGAGCA 1620
2078 TGTGTTGGAGCCTCCCGCTGAGTACCACTATTTATTTTACCTCTTTCCCAAACTGGAGCA 2137
1621 TTTATGCTTAGCTTGTCTCAAGATCTGTTACGTCCCTCTCCCTTCTCAATAAAGCATCTT 1680
2138 TTTATGCTTAGCTTGTCTCAAGATCTGTTACGTCCCTCTCCCTTCTCAATAAAGCATCTT 2197
1681 CAAGCTTGTAAAAAATAA 1703
2198 CAAGCTTGTAAAAAATAA 2220
RESULT 3
ABA91311
ID ABA91311 standard; DNA; 2159 BP.
XX ABA91311;
AC ABA91311;
DT 08-APR-2002 (first entry)
DE Human colon specific gene 337950.
XX Colon specific gene 337950.
KW Colon specific gene; CSG; gene; human; colon cancer; tumour;
XX metastasis; diagnosis; gene therapy; imaging; staging; vaccine; ds.
XX Homo sapiens.
XX WO200192528-A2.
XX 06-DEC-2001.
XX 29-MAY-2001; 2001WO-US17583.
XX 26-MAY-2000; 2000US-207383P.
XX (DIAD-) DIADEXUS INC.
PA Macina RA, Chen S, Pluta J, Sun Y, Recipon H;
XX WPI; 2002-083188/11.
XX Colon specific polypeptides and polynucleotides useful for diagnosing,
PT monitoring, staging and treating colon cancer -
XX Claim 1; Page 101-102; 116pp; English.
XX The present sequence is that of a colon specific gene (CSG),
CC designated LSGold gene 337950. This is 1 of 22 claimed CSGs (see

ABA91302-23) identified in the Incyte LIFESEQ Gold database using
the data mining Cancer Leads Automatic Search Package (CLASP),
which identifies highly expressed organ and cancer specific genes.
The CSG, or its encoded protein, can be used as a diagnostic marker
of colon cancer. The invention provides claimed methods of
diagnosing colon cancer, diagnosing metastasis of colon cancer,
staging colon cancer, monitoring colon cancer for the onset of
metastasis, and monitoring a change in stage of colon cancer in a
patient, by determining the level of a CSG in a sample of cells,
tissues or body fluid and comparing it with the level in a sample
corresponding sample from a healthy human control or from a sample
taken previously from the patient. Also claimed are methods
of identifying therapeutic agents for use in imaging and treating
colon cancer. Treatment may involve administering a compound
which downregulates expression or activity of a CSG. Adoptive
immunotherapy and gene therapy methods may be used. A claimed
vaccine for treating colon cancer comprises a claimed CSG.
XX
SQ Sequence 2159 BP; 528 A; 673 C; 581 G; 377 T; 0 other;
Query Match 78.0%; Score 1334.6; DB 24; Length 2159;
Best Local Similarity 92.0%; Pred. No. 0;
Matches 1580; Conservative 0; Mismatches 29; Indels 108; Gaps 12;
QY 1 GGCAGAGCGACTGAAGACAGCGCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGACC 60
DB 532 GGCAGAGCGACTGAAGACAGCGCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGACC 591
QY 61 TCGACTTGGAGGCGCTTACGCCAGGCCAGGACAGATGGAGGGGGCTCTGTATGGAAGGCC 120
DB 592 TCGACTTGGAGGCGCTTACGCCAGGCCAGGACAGATGGAGGGGGCTCTGTATGGAAGGCC 651
QY 121 GCTCCTCTATGAGCAGCGACGCTATCTGGAGCGGGGATCCCTCCAGAACACGCCACCA 180
DB 652 GCTCCTCTATGAGCAGCGACGCTATCTGGAGCGGGGATCCCTCCAGAACACGCCACCA 711
QY 181 GAGGACCTTAGACAGCAGCCTCCACCACATCCCAAGGGCCCTGCCACGCCACACAGTGC 240
DB 712 GAGGACCTTAGACAGCAGCCTCCACCACATCCCAAGGGCCCTGCCACGCCACACAGTGC 771
QY 241 CCGAGAACCAAGTGCCTTTACTCTGCCCTCTCCAGAGGGGCTCTCTTCCCGGAGGACC 300
DB 772 CCGAGAACCAAGTGCCTTTACTCTGCCCTCTCCAGAGGGGCTCTCTTCCCGGAGGACC 831
QY 301 AGAGAGGAGCAGGAGGAGTGTCAACCATGTCTTAAGGACATTTAGCTGTTCATGGGAAA 360
DB 832 AGAGAGGAGCAGGAGGAGTGTCAACCATGTCTTAAGGACATTTAGCTGTTCATGGGAAA 891
QY 361 GCTGGAGAGGCCCCAGGCAAAACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 417
DB 892 GCTGGAGAGGCCCCAGGCAAAACAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 951
QY 418 GGACGAGGAGGCTCTACCCAGGCCACAGTACA-TTGTACTGTCTTCAGAGAGATCAAGTACA 476
DB 952 GGACGAGGAGGCTCTACCCAGGCCACAGTACAGTGTCTTCAGAGAGATCAAGTACA 1011
QY 477 GCTTCAACCTCTCTGGAGAGGCTGGCCACCTGGCTCAAGGAGACAGTGCCTCTGAGCTCG 536
DB 1012 GCTTCAACCTCTCTGGAGAGGCTGGCCACCTGGCTCAAGGAGACAGTGCCTCTGAGCTCG 1071
QY 537 TACACATCTCTTCAAGTCCCTTGAACCTTTCATCTCGGCCAGGTGCCCTGAGGCTGCGCTAG 596
DB 1072 TACACATCTCTTCAAGTCCCTTGAACCTTTCATCTCGGCCAGGTGCCCTGAGGCTGCGCTAG 1131
QY 597 CAGCCCAAGTGTATCTACCCCTCTCCCTTCAACCTTAAAGCTATCAACCTGTACAGTCTCTGTC 656
DB 1132 CAGCCCAAGTGTATCTACCCCTCTCCCTTCAACCTTAAAGCTATCAACCTGTACAGTCTCTGTC 1191
QY 657 TAAGCCCACTGTAGAGTAACTTTTGGATGGGTTTGGGCCAGCTGGACCACTAGCCGGG 716
DB 1192 TAAGCTACCTGAGAGTAACTTTTGGATGGGTTTGGGCCAGCTGGACCACTAGCCGGG 1251
QY 717 CCGACTGGACAGGCGATGAGCCCTGCCCTACCAACCCACATTTCTCAGATGACTGGCAAC 776

Db 1252 CCGACTGGACAGGGGATGAGCCCTGCGCTACCAACCCACATTTCTCAGATGACTGGCAAC 1311
Qy 777 TTCAGAGCCCTCCAGCCAGCAGCCCTTAGGATACAGGACCCCTGTTTCCTTCGCGGG 836
Db 1312 TTCAGAGCCCTCCAGCCAGCAGCCCTTAGGATACAGGACCCCTGTTTCCTTCG 1366
Qy 837 GAAGTCATAGGTTAGGGAGCACCTCACATTTCTCTCAGGAGAACACACAACCATGACC 896
Db 1367 ----- 1366
Qy 897 CTCAGCCTGGGACCCCACTCCAGGCCCTCCAGCCCAACCTGCCCCAGCCAG-CCCTG 955
Db 1367 -----GGCTCCAGTCCCAACCTGCCCCAGCCAGTCCCTG 1402
Qy 956 AAAATGCAAGTCTGTACGAGTTTGAAGCTAGGAA--CCACAGGGAACCTCACTGTGTGTC 1013
Db 1403 AAAATGCAAGTCTGTACGAGTTTGAAGCTAGGAACTCCACGGGAACCTCACTGTGTGTC 1462
Qy 1014 AGGAGAGAAAGCTGGAGGTTCTTGACCAACAGCAGCGGTGGTGGCTGGTGAAGAAATGAGG 1073
Db 1463 AGGTAGAGAAAGCTGGAGGTTCTGACCAACAGCAGCGGTGGTGGCTGGTGAAGAAATGAGG 1522
Qy 1074 CGGACCGAGCGGCTACATTTCCAAAGCAACATCTGAGACCCCTACAGCCGGGGACCCCTG 1133
Db 1523 CGGACCGAGCGGCTACATTTCCAAAGCAACATCTGAGACCCCTACAGCCGGGGACCCCTG 1582
Qy 1134 GGACCCAGGCGCAGTCA-CCCTCTCGGGTTCATGCTTCCAGCTTACCTAGCTCGAGCCCTGAA 1192
Db 1583 GGACCCAGGCGCAGTCAACCCCTCTCGGGTTCATGCTTCCAGCTTACCTAGCTCGAGCCCTGAA 1642
Qy 1193 GAGTCAACAGCTGGCTGAGGAGGAGAACTTCTCACTCCAGCGGTGAGGACACTTGGG 1252
Db 1643 GAGTCAACAGCTGGCTGAGGAGGAGAACTTCTCACTCCAGCGGTGAGGACACTTGGG 1702
Qy 1253 TCCCTGAGGGG--AGCCAGCTACTTCGCA-TAAGACCTGGGAGCTAC--AGATGCTAT 1307
Db 1703 TCCCTGAGGGGAGCCCGCAGCTACTTCGCAATTAAGACCTGGGAGCTTACGAGGATGCTAT 1762
Qy 1308 GTCCACAGGAGGCCCCAC---GAATCCTGTCGGCTGAGGCTGTGAGAGGATGC-TG 1363
Db 1763 GTCCACAGGAGGCCCCAGCAATCTGTCGGCTGAGGCTGTGAGAGGATGCTG 1822
Qy 1364 GGGATAAGCCCTTAGGCACACAGCTTAGACACCTCCAAAGAACCCGCTGATGCAAG 1423
Db 1823 GGGATAAGCCCTTAGGCACACAGCTTAGACACCTCCAAAGAACCCGCTGATGCAAG 1882
Qy 1424 ATGCGAGATCTGATACCCATAGAGCCCCGAGAAATTCCTCTTCTGGATCCAGTTTGCAG 1483
Db 1883 ATGCGAGATCTGATACCCATAGAGCCCCGAGAAATTCCTCTTCTGGATCCAGTTTGCAG 1942
Qy 1484 CAAACCCACAC--CCCAGCTCACACAGCAAAACAAATGGACAGGCCAGAGCTGAAGC 1541
Db 1943 CAAACCCACACCTCCAGCTCACACAGCAAAACAAATGGACAGGCCAGAGCTGAAGC 2002
Qy 1542 AAACAGTGTCCCTTCTGGCTGTGTGGAGCTCCCAAGTAAACCACTATTTATTTACCT 1601
Db 2003 AAACAGTGTCCCTTCTGGCTGTGTGGAGCTTCCCAAGTAAACCACTATTTATTTACCT 2062
Qy 1602 CTTTCCCAAACTGGAGCATTTATGCTTAGCTAGGCTGTGCAAGAACTGTTAGTCCCTCTCC 1661
Db 2063 CTTTCCCAAACTGGAGCATTTATGCTTAGGCTGTGCAAGAACTGTTAGTCCCTCTCC 2122
Qy 1662 TTCTCAATAAAGCATCTTCAAGCTTCTAAAAAAA 1698
Db 2123 TTCTCAATAAAGCATCTTCAAGCTTGAATAAAAA 2159

RESULT 4
ABK43674
ID ABK43674 standard; cdna; 893 BP.
XX
AC ABK43674;

XX
DT 05-JUN-2002 (first entry)
XX
DE DNA encoding novel central nervous system protein #254.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;
KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy; gene; ss.
XX
OS Homo sapiens.
XX
PN W0200155318-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01332.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 23-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 14-SEP-2000; 2000US-0233066.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0236803.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249246.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.

PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
PI Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-581633/65.
DR P-PSDB; AAU87344.
XX
XX New isolated nucleic acid encoding a protein for diagnosing,
PT preventing, treating or ameliorating medical conditions and used as
PT food additives or preservatives -
XX
PS Claim 1; SEQ ID NO 264; 837pp; English.
XX
CC The invention describes an isolated nucleic acid molecule (I) encoding a
CC novel central nervous system protein. (I) and polypeptides (III) encoded
CC by (I), are used to treat a medical conditions and in diagnosis of a
CC pathological condition. Disorders which are diagnosed or treated include
CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative
CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders
CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,
CC angiogenesis, nervous system disorders e.g. Alzheimer's disease and
CC amyotrophic lateral sclerosis, infections caused by bacteria, viruses
CC e.g. Acquired immunodeficiency virus (AIDS) and fungi, ocular disorders
CC e.g. corneal infection, gastrointestinal disorders e.g. dysphagia,
CC adenocarcinomas and irritable bowel syndrome, reproductive system
CC disorders e.g. testicular feminisation, endocrine disorders e.g. diabetes
CC and pituitary dwarfism, cancers and disorders at the cellular level e.g.
CC leukaemia, disorders involving neovascularisation e.g. malignancies,
CC respiratory disorders e.g. nonallergic rhinitis, renal disorders e.g.
CC acute kidney failure and blood related disorders e.g. myocardial
CC infarction. The polypeptides can also be used to aid wound healing and
CC epithelial cell proliferation, to prevent skin aging due to sunburn, to
CC maintain organs before transplantation, for supporting cell culture of
CC primary tissues, to regenerate tissues and in chemotaxis. The
CC polypeptides can also be used as a food additive or preservative to
CC increase or decrease storage capabilities, fat content, lipid, protein,
Query Match 48.7%; Score 832.8; DB 23; Length 893;
Best Local Similarity 99.1%; Pred. No. 9.2e-231;
Matches 837; Conservative 0; Mismatches 8; Indels 0; Gaps 0;
Qy 866 TTTCTTCAGGAGAGACACACACATGACCTCAGCTGGGGACCCCACTCCAGGCC 925
Db 48 TTCCCTTCGGAGAGACACACACATGACCTCAGCTGGGGACCCCACTCCAGGCC 107
Qy 926 TCCAGCCCCAACTGCCAGCCAGCCCTGAATGCAAGTCTTGTAGAGTTGAAGCT 985
Db 108 TCCAGCCCCAACTGCCAGCCAGCCCTGAATGCAAGTCTTGTAGAGTTGAAGCT 167
Qy 986 AGGAACCCACGGGAACCTGACTGTGTCCAGGAGAGAGCTGGAGTTCTGACACACAGC 1045
Db 168 AGGAACCCACGGGAACCTGACTGTGTCCAGGAGAGAGCTGGAGTTCTGACACACAGC 227
Qy 1046 AAGCGGTGGTGGCTGGTGAAGAATGAGCGGGAGCGGCTACATTCCAAAGCAATC 1105
Db 228 AAGCGGTGGTGGCTGGTGAAGAATGAGCGGGAGCGGCTACATTCCAAAGCAATC 287
Qy 1106 CTGAGGCCCTTACAGCCGGGGACCCCTGGGACCCAGGCCAGTCACCTCTCGGGTTCCA 1165

|||||
Db 288 CTGGAGCCCTACAGCCGGGACCCCTGGGACCCAGGCCAGTCACTCTCGGGTTCCA 347
Qy 1166 ATGCTTCGACTTACTCGAGCCCTGAAGAGGTCACAGACTGGCTGCAGGAGAGAACTTC 1225
Db 348 ATGCTTCGACTTACTCGAGCCCTGAAGAGGTCACAGACTGGCTGCAGGAGAGAACTTC 407
Qy 1226 TCCACTGCCAGGTGAGGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGA 1285
Db 408 TCCACTGCCAGGTGAGGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGA 467
Qy 1286 CCTGGGAGCTACAGATGCTATGTCCACAGAGGCCGCCAGAAATCCCTGTCGCCGCTGGAG 1345
Db 468 CCTGGGAGCTACAGATGCTATGTCCACAGAGGGGCCACGAAATCCCTGTCGCCGCTGGAG 527
Qy 1346 GCTGTCAGAGGATGCTGGGATAGCCCTTAGGCACACAGCTTAGACACCTCCAGAAC 1405
Db 528 GCTGTCAGAGGATGCTGGGATAGCCCTTAGGCACACAGCTTAGACACCTCCAGAAC 587
Qy 1406 AGGCCCGCTGATGCAAGATGGCAGATCTGTATACCCATTAGAGCCCGAGAAATTCCTCTT 1465
Db 588 AGGCCCGCTGATGCAAGATGGCAGATCTGTATACCCATTAGAGCCCGAGAAATTCCTCTT 647
Qy 1466 CTGATCCCGATTGTCAGCAAAACCCACAGCCAGCTACACAGCAAAACATGGACAG 1525
Db 648 CTGATCCCGATTGTCAGCAAAACCCACAGCCAGCTACACAGCAAAACATGGACAG 707
Qy 1526 GCCCAGAGGCTGAGCAAAACAGTCTCCCTCTGCTGTGTGGAGCCCTCCCAAGTACCA 1585
Db 708 GCCCAGAGGCTGAGCAAAACAGTCTCCCTCTGCTGTGTGGAGCCCTCCCAAGTACCA 767
Qy 1586 CCTATTATTTTACCTCTTTCCCAACCTGGAGCATTTATGCCCTAGGCTTGTCAAGAAATC 1645
Db 768 CCTATTATTTTACCTCTTTCCCAACCTGGAGCATTTATGCCCTAGGCTTGTCAAGAAATC 827
Qy 1646 TGTTCAGTCCCTCTCTCAATAAAGACATCTTCAAGCTTGTAAATAAATAAANG 1705
Db 828 TGTTCAGTCCCTCTCTCAATAAAGCATCTTCAAGCTTGTAAATAAATAAATAAANG 887
Qy 1706 ATAAA 1710
Db 888 AAAAA 892
RESULT 5
AAH33515
ID AAH33515 standard; cDNA; 872 BP.
AC AAH33515;
XX 03-SEP-2001 (first entry)
DT Human colon cancer antigen encoding cDNA SEQ ID NO:571.
DE Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW colorectal carcinoma; ss.
XX Homo sapiens.
OS
XX WO200122920-A2.
PN
XX 05-APR-2001.
PD
XX 28-SEP-2000; 2000WO-US26524.
PF
XX 29-SEP-1999; 99US-0157137.
PR 03-NOV-1999; 99US-0163280.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
PI Ruben SM, Barash SC, Birse CE, Rosen CA;
XX WPI; 2001-235357/24.
DR

DR P-PSDB; AAG74084.
XX Nucleic acids encoding 4277 human colon cancer-associated polypeptides,
PT useful for preventing, diagnosing and/or treating colorectal cancers -
XX Claim 1; Page 2629; 9803pp; English.
XX AAH32943 to AAH37195 and AAG73514 to AAG77788 represent human colon
CC cancer-associated nucleic acid molecules (N) and proteins (P), where
CC the proteins are collectively known as colon cancer antigens. The colon
CC cancer antigens have cytostatic activity and can be used in gene
CC therapy and vaccine production. N and P may be used in the prevention,
CC diagnosis and treatment of diseases associated with inappropriate P
CC expression. For example, N and P may be used to treat disorders
CC associated with decreased expression by rectifying mutations or deletions
CC in a patient's genome that affect the activity of P by expressing
CC inactive proteins or to supplement the patients own production of P.
CC Additionally, N may be used to produce the colon cancer-associated Ps,
CC by inserting the nucleic acids into a host cell and culturing the cell
CC to express the proteins. N and P can be used in the prevention, diagnosis
CC and treatment of colorectal carcinomas and cancers. AAH37196 to AAH37204
CC and AAB77789 represent sequences used in the exemplification of the
CC present invention.
CC N.B. Pages 666 to 682 and page 7053 of the sequence listing were
CC missing at time of publication, meaning no sequences are present for
CC SEQ ID NO:1027 to 1052, 7921 and 7922.
XX
SQ Sequence 872 BP; 225 A; 261 C; 218 G; 164 T; 4 other;
Query Match 48.6%; Score 830.8; DB 22; Length 872;
Best Local Similarity 99.4%; Pred. No. 3.5e-230;
Matches 832; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
Qy 866 TTTCTCTCAGAGAGACACACACCATGACCCTCAGCTGGGGACCCCACTCCAGGCC 925
Db 30 TTCCCTTCGGAGAGACACACACCATGACCCTCAGCTGGGGACCCCACTCCAGGCC 89
Qy 926 TCCAGCCCCAAACCTGCCGCCAGCCCTGAAAATGCAAGTCTTGTACGAGTTTGAAGCT 985
Db 90 TCCAGCCCCAAACCTGCCGCCAGCCCTGAAAATGCAAGTCTTGTACGAGTTTGAAGCT 149
Qy 986 AGGAACCCAGGGAACTGACTGTGGTCCAGGGAGAGAGCTGGAGTTCTGGACACAGC 1045
Db 150 AGGAACCCAGGGAACTGACTGTGGTCCAGGGAGAGAGCTGGAGTTCTGGACACAGC 209
Qy 1046 AAGCGTGTGGCTGTGAAGATGAGCGGGAGCGGCTACATTCCAAGCAACATC 1105
Db 210 AAGCGTGTGGCTGTGAAGATGAGCGGGAGCGGCTACATTCCAAGCAACATC 269
Qy 1106 CTGGAGCCCTTACAGCCGGGACCCCTGGGACCCAGGCCAGTCACTCTCGGGTTCCA 1165
Db 270 CTGGAGCCCTTACAGCCGGGACCCCTGGGACCCAGGCCAGTCACTCTCGGGTTCCA 329
Qy 1166 ATGCTTCGACTTAGCTCAGGCCCTGAAGAGTCAAGACTGGCTGCAGGAGAGAACTTC 1225
Db 330 ATGCTTCGACTTAGCTCAGGCCCTGAAGAGTCAAGACTGGCTGCAGGAGAGAACTTC 389
Qy 1226 TCCACTGCCAGGTGAGGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGA 1285
Db 390 TCCACTGCCAGGTGAGGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGA 449
Qy 1286 CTGGGAGGCTACAGATGCTATGTCCACAGAGGCCCCACGAATCTCTCCGGCTGGAG 1345
Db 450 CTGGGAGGCTACAGATGCTATGTCCACAGAGGCCCCACGAATCTCTCCGGCTGGAG 509
Qy 1346 GCTGTCAGAGGATGCTGGGATAGCCCTTAGGCACACAGCTTAGACACCTCCAGAAC 1405
Db 510 GCTGTCAGAGGATGCTGGGATAGCCCTTAGGCACACAGCTTAGACACCTCCAGAAC 569
Qy 1406 AGGCCCGCTGATGCAAGATGGCAGATCTGATACCCATTAGAGCCCGAGAAATTCCTCTT 1465
Db 570 AGGCCCGCTGATGCAAGATGGCAGATCTGATACCCATTAGAGCCCGAGAAATTCCTCTT 629

QY	1466	CTGGATCCAGTTTCAGCAAAACCCACACCCAGCTCACACAGCAAAACAATGGACAG	1525	PR	14-AUG-2000;	2000US-0225757.
Db	630	CTGGATCCAGTTTCAGCAAAACCCACACCCAGCTCACACAGCAAAACAATGGACAG	689	PR	14-AUG-2000;	2000US-0225758.
QY	1526	GCCAGAGGCTGAAGCAACAGTGCCTTCTGGCTGTGTTGGAGCCTCCCGAGTAACCA	1585	PR	18-AUG-2000;	2000US-0226279.
Db	690	GCCAGAGGCTGAAGCAACAGTGCCTTCTGGCTGTGTTGGAGCCTCCCGAGTAACCA	749	PR	22-AUG-2000;	2000US-0226681.
QY	1586	CTATTATTATTTACCTCTTCCCAAACTGGAGCATTTATGCCTAGGCTTCTCAAGAATC	1645	PR	22-AUG-2000;	2000US-0226868.
Db	750	CTATTATTATTTACCTCTTCCCAAACTGGAGCATTTATGCCTAGGCTTCTCAAGAATC	809	PR	23-AUG-2000;	2000US-0227182.
QY	1646	TGTTTCAGTCCCTCTCTTCTCAATAAAAGCATCTTCAAGCTGTGTAATAAAAAATA	1702	PR	30-AUG-2000;	2000US-0227009.
Db	810	TGTTTCAGTCCCTCTCTTCTCAATAAAAGCATCTTCAAGCTGTGTAATAAAAAANA	866	PR	01-SEP-2000;	2000US-0228924.
RESULT 6				PR	01-SEP-2000;	2000US-0229287.
AAS27109				PR	01-SEP-2000;	2000US-0229343.
XX				PR	01-SEP-2000;	2000US-0229344.
AC				PR	01-SEP-2000;	2000US-0229345.
XX				PR	05-SEP-2000;	2000US-0229509.
DT	07-NOV-2001	(first entry)		PR	05-SEP-2000;	2000US-0229513.
DE		cdna encoding novel signal transduction pathway protein, Seq ID 144.		PR	06-SEP-2000;	2000US-0230437.
KW		Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;		PR	06-SEP-2000;	2000US-0230438.
KW		antiflammatory; anti-HIV; antibacterial; antineoplastic; antineoplastic; cancer;		PR	08-SEP-2000;	2000US-0231242.
KW		immune system disorder; rheumatoid arthritis; inflammatory condition;		PR	08-SEP-2000;	2000US-0231243.
KW		organ transplant rejection; infection; hepatitis C; blood disorder;		PR	08-SEP-2000;	2000US-0231244.
KW		sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;		PR	08-SEP-2000;	2000US-0231413.
KW		neurodegenerative disorder; Alzheimer's disease; parkinson's disease;		PR	08-SEP-2000;	2000US-0231414.
KW		chromosomal abnormality; Down syndrome; ischaemia; renal disorder;		PR	08-SEP-2000;	2000US-0232080.
KW		cardiovascular; respiratory; wound healing; endocrine; Addison's disease;		PR	08-SEP-2000;	2000US-0232081.
KW		reproductive system; gastrointestinal; liver disorder; AIDS; ss;		PR	12-SEP-2000;	2000US-0231968.
KW		acquired immune deficiency syndrome.		PR	14-SEP-2000;	2000US-0232397.
XX				PR	14-SEP-2000;	2000US-0232398.
OS	Homo sapiens.			PR	14-SEP-2000;	2000US-0232399.
PN	WO200154733-Al.			PR	14-SEP-2000;	2000US-0232400.
XX				PR	14-SEP-2000;	2000US-0232401.
PD	02-AUG-2001.			PR	14-SEP-2000;	2000US-0233063.
XX				PR	14-SEP-2000;	2000US-0233064.
PF	17-JAN-2001;	2001WO-US01312.		PR	14-SEP-2000;	2000US-0233065.
XX				PR	21-SEP-2000;	2000US-0234223.
PR	31-JAN-2000;	2000US-0179065.		PR	21-SEP-2000;	2000US-0234274.
PR	04-FEB-2000;	2000US-0180628.		PR	25-SEP-2000;	2000US-0234997.
PR	24-FEB-2000;	2000US-0184664.		PR	25-SEP-2000;	2000US-0234998.
PR	02-MAR-2000;	2000US-0186350.		PR	26-SEP-2000;	2000US-0235484.
PR	16-MAR-2000;	2000US-0189874.		PR	27-SEP-2000;	2000US-0235834.
PR	17-MAR-2000;	2000US-0190076.		PR	27-SEP-2000;	2000US-0235836.
PR	18-APR-2000;	2000US-0198123.		PR	29-SEP-2000;	2000US-0236327.
PR	19-MAY-2000;	2000US-0205515.		PR	29-SEP-2000;	2000US-0236367.
PR	07-JUN-2000;	2000US-0209467.		PR	29-SEP-2000;	2000US-0236368.
PR	28-JUN-2000;	2000US-0214886.		PR	29-SEP-2000;	2000US-0236369.
PR	30-JUN-2000;	2000US-0215135.		PR	29-SEP-2000;	2000US-0236370.
PR	07-JUL-2000;	2000US-0216647.		PR	02-OCT-2000;	2000US-0236802.
PR	07-JUL-2000;	2000US-0216880.		PR	02-OCT-2000;	2000US-0237037.
PR	11-JUL-2000;	2000US-0217487.		PR	02-OCT-2000;	2000US-0237038.
PR	11-JUL-2000;	2000US-0217496.		PR	02-OCT-2000;	2000US-0237039.
PR	14-JUL-2000;	2000US-0218290.		PR	02-OCT-2000;	2000US-0237040.
PR	26-JUL-2000;	2000US-0220963.		PR	13-OCT-2000;	2000US-0239935.
PR	26-JUL-2000;	2000US-0220964.		PR	13-OCT-2000;	2000US-0239937.
PR	14-AUG-2000;	2000US-0224518.		PR	20-OCT-2000;	2000US-0240960.
PR	14-AUG-2000;	2000US-0224519.		PR	20-OCT-2000;	2000US-0241221.
PR	14-AUG-2000;	2000US-0225213.		PR	20-OCT-2000;	2000US-0241785.
PR	14-AUG-2000;	2000US-0225214.		PR	20-OCT-2000;	2000US-0241786.
PR	14-AUG-2000;	2000US-0225266.		PR	20-OCT-2000;	2000US-0241787.
PR	14-AUG-2000;	2000US-0225267.		PR	20-OCT-2000;	2000US-0241808.
PR	14-AUG-2000;	2000US-0225268.		PR	20-OCT-2000;	2000US-0241809.
PR	14-AUG-2000;	2000US-0225270.		PR	20-OCT-2000;	2000US-0241826.
PR	14-AUG-2000;	2000US-0225447.		PR	01-NOV-2000;	2000US-0244617.
PR				PR	08-NOV-2000;	2000US-0246474.
PR				PR	08-NOV-2000;	2000US-0246475.
PR				PR	08-NOV-2000;	2000US-0246476.
PR				PR	08-NOV-2000;	2000US-0246477.
PR				PR	08-NOV-2000;	2000US-0246478.
PR				PR	08-NOV-2000;	2000US-0246523.
PR				PR	08-NOV-2000;	2000US-0246524.
PR				PR	08-NOV-2000;	2000US-0246525.
PR				PR	08-NOV-2000;	2000US-0246526.
PR				PR	08-NOV-2000;	2000US-0246527.
PR				PR	08-NOV-2000;	2000US-0246528.
PR				PR	08-NOV-2000;	2000US-0246532.

PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249267.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256179.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-02559678.
XX
PA (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
PI
XX
DR WPI; 2001-465460/50.
DR P-PSDB; AAU17192.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders
XX
XX Claim 1; SEQ ID No 144; 880pp: English.
PS
XX The invention relates to novel isolated polypeptides (I), and
XX polynucleotides (II). (I), (II) and the antibody to (I) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in
CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
CC pathway protein coding sequences and PCR primers of the invention.
XX

Query Match 48.5%; Score 829.6; DB 22; Length 872;
Best Local Similarity 99.4%; Pred. No. 7.7e-230;
Matches 832; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 866 TTTCCTCAGGAGAGACACACACCATGAGCCCTCAGCCTGGGGACCCCACTCCAGGCC 925
Db 30 TTCCCTTCGGAGAGACACACACCATGAGCCCTCAGCCTGGGGACCCCACTCCAGGCC 89
Qy 926 TCCAGCCCCAAACCTGCCAGCCGACCTGAAATGCAAGTCTTTGTACGAGTTTGAAGCT 985
Db 90 TCCAGCCCCAAACCTGCCAGCCGACCTGAAATGCAAGTCTTTGTACGAGTTTGAAGCT 149
Qy 986 AGGAACCCACGGGAACCTGACTGTGTCAGGGAGAGAAGCTGGAGGTTCTGGACCACAGC 1045
Db 150 AGGAACCCACGGGAACCTGACTGTGTCAGGGAGAGAAGCTGGAGGTTCTGGACCACAGC 209
Qy 1046 AAGCGGTGGTGGTGTGAAGATGAGGGGGAGCGGCTACATTTCCAAGCAACATC 1105
Db 210 AAGCGGTGGTGGTGTGAAGATGAGGGGGAGCGGCTACATTTCCAAGCAACATC 269
Qy 1106 CTGGAGCCCTACAGCCGGGACCCCTGGGACCCAGGCCAGTCCCTCTCGGGTTCCA 1165
Db 270 CTGGAGCCCTACAGCCGGGACCCCTGGGACCCAGGCCAGTCCCTCTCGGGTTCCA 329
Qy 1166 ATGCTTCGACTTAGCTCGAGGCTTGAAGAGGTACAGACTGGCTGCAGGAGAGAATTC 1225
Db 330 ATGCTTCGACTTAGCTCGAGGCTTGAAGAGGTACAGACTGGCTGCAGGAGAGAATTC 389
Qy 1226 TCCACTGCCACGGTGGAGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGA 1285
Db 390 TCCACTGCCACGGTGGAGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGA 449
Qy 1286 CCTGGGGAGCTACAGATGCTATGTCCACAGGAGGCCCCACGAATCTCTCCCGCTGGAG 1345
Db 450 CCTGGGGAGCTACAGATGCTATGTCCACAGGAGGCCCCACGAATCTCTCCCGCTGGAG 509
Qy 1346 GCTGTCAAGAGGATGCTGGGATAAGCCCTTAGGCACACAGCTTAGACACCTTCCAAGAACC 1405
Db 510 GCTGTCAAGAGGATGCTGGGATAAGCCCTTAGGCACACAGCTTAGACACCTTCCAAGAACC 569
Qy 1406 AGGCCCCGCTGATGCAAGATGGCAGATCTGATACCCATTAGAGCCCCGAGAAATTCCTCTT 1465
Db 570 AGGCCCCGCTGATGCAAGATGGCAGATCTGATACCCATTAGAGCCCCGAGAAATTCCTCTT 629
Qy 1466 CTGGATCCCAGTTGCGAGCAAAACCCACAGCCAGCTACACACCAAAACAATGGACAG 1525
Db 630 CTGGATCCCAGTTGCGAGCAAAACCCACAGCCAGCTACACACCAAAACAATGGACAG 689
Qy 1526 GCCCAGAGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGAGCTTCCCGAGTAACCA 1585
Db 690 GCCCAGAGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGAGCTTCCCGAGTAACCA 749
Qy 1586 CCTATTATTTTACCTCTTTCCCAAACTGGAGCAATTTATGCCTTAGCTTGTCAAGAATC 1645
Db 750 CCTATTATTTTACCTCTTTCCCAAACTGGAGCAATTTATGCCTTAGCTTGTCAAGAATC 809
Qy 1646 TGTTCACTCCCTCTCTCTCTCAATAAAGATCTTCAAGCTTGTAAAAAANAATA 1702
Db 810 TGTTCACTCCCTCTCTCTCTCAATAAAGATCTTCAAGCTTGTAAAAAANAATA 866

RESULT 7
ABK43957
ID ABK43957 standard; cDNA; 872 BP.
XX
AC ABK43957;
XX
DT 05-JUN-2002 (first entry)
XX
DE DNA encoding novel central nervous system protein #537.
XX
KW Central nervous system; CNS; autoimmune disease; rheumatoid arthritis;

KW hyperproliferative disorder; neoplasm; cardiovascular disorder;
KW cardiac arrest; cerebrovascular disorder; ischaemia; anglogenesis;
KW nervous system disorder; Alzheimer's disease; AIDS; ocular disorder;
KW acquired immunodeficiency virus; dysphagia; gastrointestinal disorder;
KW adenocarcinoma; reproductive system disorder; testicular feminisation;
KW endocrine disorder; diabetes; cancer; leukaemia; neovascularisation;
KW respiratory disorder; renal disorder; kidney failure; blood disorder;
KW myocardial infarction; wound healing; cell proliferation; skin aging;
KW food additive; food preservative; gene therapy; gene; ss.
OS Homo sapiens.
XX
PN WO200155318-A2.
XX
PD
PF 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01332.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 12-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.

PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226688.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236368.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246479.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI; 2001-465460/50.
XX P-PSDB; AAU17191.
XX
XX Novel polypeptides useful for diagnosing, treating, preventing and/or
XX prognosing disorders related to the proteins, including cancers, immune
XX disorders and neuronal disorders
XX
XX Claim 1; SEQ ID No 143; 880pp; English.
XX
XX The invention relates to novel isolated polypeptides (I), and
XX polynucleotides (II). (I), (II) and the antibody to (I) are useful for
XX diagnosing, preventing and treating diseases including immune system
XX disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
XX disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
XX transplant rejections and graft versus host disease, infectious diseases

RESULT 9
ABK43673

Db 783 GCTGCAGGACAGAACTTCCACTGCCACGGTCAGGACACTTGGGTCCCTGACGGGAG 842
QY 1267 CCAGCTACTTCGGATAGACCTGGGAGCTACAGATGCTATGTCCACAGAGGCCCCACG 1326
Db 843 CCAGCTACTTCGGATAGACCTGGGAGCTACAGATGCTATGTCCACAGAGGCCCCACG 902
QY 1327 AATCCTGTCGCCGCTGGAGGCTGTCAAGAGATGCTGGGATAAGCCCTTAGGCACG 1386
Db 903 AATCCTGTCGCCGCTGGAGGCTGTCAAGAGATGCTGGGATAAGCCCTTAGGCACG 962
QY 1387 TTAGACACCTCCAGACACGAGCCCGCTGATGCAAGATGGCAGATGATATCCCATTTAG 1446
Db 963 TTAGACACCTCCAGACACGAGCCCGCTGATGCAAGATGGCAGATGATATCCCATTTAG 1022
QY 1447 AGCCCCGAGAAATTCCTCTTCTGGATCCAGTTCGACGAAACCCACACCCAGCTCACA 1506
Db 1023 AGCCCCGAGAAATTCCTCTTCTGGATCCAGTTCGACGAAACCCACACCCAGCTCACA 1082
QY 1507 CAGCAAAAACATGGACAGGCCAGAGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTT 1566
Db 1083 CAGCAAAAACATGGACAGGCCAGAGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTT 1142
QY 1567 GGAGCCTCCCGCAGTAACCACTATTTATTTTACCTCTTTCCCAACCTGGAGCATTTATG 1626
Db 1143 GGAGCCTCCCGCAGTAACCACTATTTATTTTACCTCTTTCCCAACCTGGAGCATTTATG 1202
QY 1627 CCTAGGCTTCTCAAGAACTGTTCAGTCCCTCTCTCTCAATAAAAGCATCTTCAAGCT 1686
Db 1203 CCTAGGCTTCTCAAGAACTGTTCAGTCCCTCTCTCTCAATAAAAGCATCTTCAAGCT 1262
QY 1687 TGTAAAAAAAATAANGA 1706
Db 1263 TGTAAAAAAAATAANGA 1282

RESULT 10
ABN93792/c
ID ABN93792 standard; DNA; 401 BP.
XX
AC ABN93792;
XX
DT 13-AUG-2002 (first entry)
XX
XX Gene #290 used to diagnose liver cancer.
KW Gene; liver cancer; ds; hepatocellular carcinoma; hepatotrophic;
KW metastatic liver tumor; cytostatic; expression profile; disease state;
KW disease progression; drug toxicity; drug efficacy; drug metabolism.
XX
OS Homo sapiens.
XX
PN WO200229103-A2.
XX
PD 11-APR-2002.
XX
XX 02-OCT-2001; 2001WO-US30589.
XX
XX 02-OCT-2000; 2000US-237054P.
XX
XX (GENE-) GENE LOGIC INC.
XX
XX Horne D, Alvares C, Peres-Da-Silva S, Vockley JG;
PI WPI; 2002-426119/45.
XX
XX Diagnosing and detecting the progression of liver cancer,
PT hepatocellular carcinoma or metastatic liver tumor in a patient,
PT involves detecting the level of expression of two or more genes in a
PT liver tissue sample -
XX
XX Claim 1; SEQ ID NO 290; 298pp; English.
XX
XX The invention relates to a novel method for diagnosing and detecting the

CC progression of liver cancer, hepatocellular carcinoma or metastatic liver
CC tumour in a patient, and differentiating metastatic liver cancer from
CC hepatocellular carcinoma in a patient, involving detecting the level of
CC expression of two or more genes represented in ABN93503-ABN97455 in a
CC tissue sample. The method of the invention has hepatotropic, and
CC cytostatic activity. The method is useful for diagnosing and detecting
CC the progression of liver cancer, hepatocellular carcinoma and metastatic
CC liver carcinoma in a patient. The method is useful for identifying
CC expression profiles which serve as useful diagnostic markers as well as
CC markers that can be used to monitor disease states, disease progression,
CC drug toxicity, drug efficacy and drug metabolism.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 401 BP; 83 A; 93 C; 124 G; 101 T; 0 other;

Query Match 22.7%; Score 389; DB 24; Length 401;
Best Local Similarity 99.8%; Pred. No. 2.6e-102;
Matches 400; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1259 ACGGGAGCCAGCTACTTCGCATAAGACCTGGGGAGCTACAGATGCTATGTCCA-CAGGA 1317
Db 401 ACGGGAGCCAGCTACTTCGCATAAGACCTGGGGAGCTACAGATGCTATGTCCACCAGGA 342
QY 1318 GGCCCCCAGCAATCTCTCCCGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCTTA 1377
Db 341 GGCCCCCAGCAATCTCTCCCGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCTTA 282
QY 1378 GGCAACAGCTTACAGACCTCCAAAGAACCCAGGCCCTGTATGCAAGATGGCAGATCTGAT 1437
Db 281 GGCAACAGCTTACAGACCTCCAAAGAACCCAGGCCCTGTATGCAAGATGGCAGATCTGAT 222
QY 1438 ACCATTAGAGCCCCGAGAAATTCCTCTTCTGGATCCAGTTCGACAAACCCACACCC 1497
Db 221 ACCATTAGAGCCCCGAGAAATTCCTCTTCTGGATCCAGTTCGACAAACCCACACCC 162
QY 1498 CAGCTCACACAGCAAAAACAAATGACAGGCCAGAGGCTGAAGCAACACAGTGTCCCTTCT 1557
Db 161 CAGCTCACACAGCAAAAACAAATGACAGGCCAGAGGCTGAAGCAACACAGTGTCCCTTCT 102
QY 1558 GGCTGTGTGGAGCCCTCCCGCAGTAACCACTATTTATTTTACCTCTTTTCCCAACCTGGA 1617
Db 101 GGCTGTGTGGAGCCCTCCCGCAGTAACCACTATTTATTTTACCTCTTTTCCCAACCTGGA 42
QY 1618 GCATTTATGCTAGGCTTGTCAAGAAATCTGTTTCCAGTCCCTC 1658
Db 41 GCATTTATGCTAGGCTTGTCAAGAAATCTGTTTCCAGTCCCTC 1

RESULT 11
AAK88048
ID AAK88048 standard; cDNA; 1388 BP.
XX
AC AAK88048;
XX
DT 05-NOV-2001 (first entry)
XX
DE Human digestive system antigen coding sequence SEQ ID NO: 364.
XX
KW Human; digestive system antigen; gene therapy; cancer; appendicitis;
KW ulcerative colitis; infection; Hirschsprung's disease; chronic colitis;
KW digestive system disorder; Meckel's diverticulum; ss.
XX
OS Homo sapiens.
XX
PN WO200155314-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001WO-US01324.
XX
PR 31-JAN-2000; 2000US-0179065.

```
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.
PR 14-AUG-2000; 2000US-0225447.
PR 14-AUG-2000; 2000US-0225757.
PR 14-AUG-2000; 2000US-0225758.
PR 14-AUG-2000; 2000US-0225759.
PR 18-AUG-2000; 2000US-0226279.
PR 22-AUG-2000; 2000US-0226681.
PR 22-AUG-2000; 2000US-0226686.
PR 22-AUG-2000; 2000US-0227182.
PR 23-AUG-2000; 2000US-0227009.
PR 30-AUG-2000; 2000US-0228924.
PR 01-SEP-2000; 2000US-0229287.
PR 01-SEP-2000; 2000US-0229343.
PR 01-SEP-2000; 2000US-0229344.
PR 01-SEP-2000; 2000US-0229345.
PR 05-SEP-2000; 2000US-0229509.
PR 05-SEP-2000; 2000US-0229513.
PR 06-SEP-2000; 2000US-0230437.
PR 06-SEP-2000; 2000US-0230438.
PR 08-SEP-2000; 2000US-0231242.
PR 08-SEP-2000; 2000US-0231243.
PR 08-SEP-2000; 2000US-0231244.
PR 08-SEP-2000; 2000US-0231413.
PR 08-SEP-2000; 2000US-0231414.
PR 08-SEP-2000; 2000US-0232080.
PR 08-SEP-2000; 2000US-0232081.
PR 12-SEP-2000; 2000US-0231968.
PR 14-SEP-2000; 2000US-0232397.
PR 14-SEP-2000; 2000US-0232398.
PR 14-SEP-2000; 2000US-0232399.
PR 14-SEP-2000; 2000US-0232400.
PR 14-SEP-2000; 2000US-0232401.
PR 14-SEP-2000; 2000US-0233063.
PR 14-SEP-2000; 2000US-0233064.
PR 14-SEP-2000; 2000US-0233065.
PR 21-SEP-2000; 2000US-0234223.
PR 21-SEP-2000; 2000US-0234274.
PR 25-SEP-2000; 2000US-0234997.
PR 25-SEP-2000; 2000US-0234998.
PR 26-SEP-2000; 2000US-0235484.
PR 27-SEP-2000; 2000US-0235834.
PR 27-SEP-2000; 2000US-0235836.
PR 29-SEP-2000; 2000US-0236327.
PR 29-SEP-2000; 2000US-0236367.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236369.
PR 29-SEP-2000; 2000US-0236370.
PR 02-OCT-2000; 2000US-0236802.
PR 02-OCT-2000; 2000US-0237037.
PR 02-OCT-2000; 2000US-0237038.
PR 02-OCT-2000; 2000US-0237039.
PR 02-OCT-2000; 2000US-0237040.
PR 13-OCT-2000; 2000US-0239935.
PR 13-OCT-2000; 2000US-0239937.
PR 20-OCT-2000; 2000US-0240960.
PR 20-OCT-2000; 2000US-0241221.
PR 20-OCT-2000; 2000US-0241785.
PR 20-OCT-2000; 2000US-0241786.
PR 20-OCT-2000; 2000US-0241787.
PR 20-OCT-2000; 2000US-0241808.
PR 20-OCT-2000; 2000US-0241809.
PR 20-OCT-2000; 2000US-0241826.
PR 01-NOV-2000; 2000US-0244617.
PR 08-NOV-2000; 2000US-0246474.
PR 08-NOV-2000; 2000US-0246475.
PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 17-NOV-2000; 2000US-0249300.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 05-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 08-DEC-2000; 2000US-0251990.
PR 11-DEC-2000; 2000US-0254097.
PR 05-JAN-2001; 2001US-0259678.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Barash SC, Ruben SM;
XX
XX WPI: 2001-502630/55.
XX P-PSDB: AAM92275.
XX
XX Polynucleotides encoding digestive system antigens, useful for
XX diagnosing, treating, preventing and/or prognosing disorders of the
XX digestive system, particularly cancer and cancer metastases -
XX PT
```

XX PS Claim 1; SEQ ID NO 364; 986pp; English.

XX CC The present invention provides the protein and coding sequences of a

CC number of human digestive system antigens. These can be used in the

CC diagnosis, treatment and prevention of digestive system disorders,

CC including cancer, Meckel's diverticulum, bacterial or parasitic

CC infections, appendicitis, Hirschsprung's disease, chronic colitis or

CC ulcerative colitis. The present sequence is a cDNA encoding a digestive

CC system antigen of the invention.

XX SQ Sequence 1388 BP; 329 A; 432 C; 377 G; 238 T; 12 other;

Query Match 22.7%; Score 388.8; DB 22; Length 1388;

Best Local Similarity 97.5%; Pred. No. 5.7e-102;

Matches 422; Conservative 3; Mismatches 3; Indels 5; Gaps 3;

QY 1 GGCAGAGCGACTGAAGCACCAGCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAA---AG 57

Db 615 GGCAGAGCGACTGAAGCACCAGCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAAGCAG 674

QY 58 ACCTCGACTTGGAGGCTTCAGCCAGGCCAGGACAGATGGAGGGGCGCTGTATGGAAG 117

Db 675 ACCTCGACTTGGAGGCTTCAGCCAGGCCAGGACAGATGGAGGGGCGCTGTATGGAAG 734

QY 118 GCGCGTCCCTATGGAGCAGCAGCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCCA 177

Db 735 GCGCGTCCCTATGGAGCAGCAGCTATCTGGAGCCGGGATCCCTTCAGAACAGCCCCA 794

QY 178 CCAGAGGACCTAGAGCAGCAGCTCCACCATCCCAAGGCCCTGCAGGCCACACACAG 237

Db 795 CCAGAGGACCTAGAGCAGCAGCTCCACCATCCCAAGGCCCTGCAGGCCACACACAG 854

QY 238 TGCCCGAGAACCAAGTGCCCTTTACTTGCTCTCTCAAGSCGCTCTTCCCGCCGAGGA 297

Db 855 TGCCCGAGAACCAAGTGCCCTTTACTTGCTCTCTCAAGSCGCTCTTCCCGCCGAGGA 914

QY 298 CCAGAGAGGAGCAGGAGAGTGTCTGAACCATGCTCTTAAGGACATTTAGCTGTTCATGG 357

Db 915 CCAGAGAGGAGCAGGAGAGTGTCTGAACCATGCTCTTAAGGACATTTGA-CTGNTCATGG 973

QY 358 AAGCTGGAGAGGCCAGCCAGCAAGCAGCAGCAGGAGAGAGAAATTTGGGAAAAAACAA 417

Db 974 AAT-CTGGANAAGCCAGGCAAGCAGCAGGAGAGAGAAATTTGGGAAAAAACAA 1032

QY 418 GGACAGGGAGGT 430

Db 1033 GGACAGGGAGGT 1045

RESULT 12

AAS27487

ID AAS27487 standard; cDNA; 572 BP.

XX AC AAS27487;

XX DF 07-NOV-2001 (first entry)

XX DE cDNA encoding novel signal transduction pathway protein, Seq ID 522.

XX KW Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour;

KW antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer;

KW immune system disorder; rheumatoid arthritis; inflammatory condition;

KW organ transplant rejection; hepatitis C; blood disorder;

KW sickle cell anaemia; hyperproliferative disorder; Gaucher's disease;

KW neurodegenerative disorder; Alzheimer's disease; Parkinson's disease;

KW chromosomal abnormality; Down syndrome; ischaemia; renal disorder;

KW cardiovascular; respiratory; wound healing; endocrine; Addison's disease;

KW reproductive system; gastrointestinal; liver disorder; AIDS; ss;

XX OS acquired immune deficiency syndrome.

OS Homo sapiens.

PN XX WO200154733-A1.

PD XX 02-AUG-2001.

XX 17-JAN-2001; 2001WO-US01312.

XX 31-JAN-2000; 2000US-0179065.

PR 04-FEB-2000; 2000US-0180628.

PR 24-FEB-2000; 2000US-0184664.

PR 02-MAR-2000; 2000US-0186350.

PR 16-MAR-2000; 2000US-0189874.

PR 17-MAR-2000; 2000US-0190076.

PR 18-APR-2000; 2000US-0198123.

PR 19-MAY-2000; 2000US-0205515.

PR 07-JUN-2000; 2000US-0209467.

PR 28-JUN-2000; 2000US-0214886.

PR 30-JUN-2000; 2000US-0215135.

PR 07-JUL-2000; 2000US-0216647.

PR 07-JUL-2000; 2000US-0216880.

PR 11-JUL-2000; 2000US-0217487.

PR 11-JUL-2000; 2000US-0217496.

PR 14-JUL-2000; 2000US-0218290.

PR 26-JUL-2000; 2000US-0220963.

PR 26-JUL-2000; 2000US-0220964.

PR 14-AUG-2000; 2000US-0224518.

PR 14-AUG-2000; 2000US-0224519.

PR 14-AUG-2000; 2000US-0225213.

PR 14-AUG-2000; 2000US-0225214.

PR 14-AUG-2000; 2000US-0225266.

PR 14-AUG-2000; 2000US-0225267.

PR 14-AUG-2000; 2000US-0225268.

PR 14-AUG-2000; 2000US-0225270.

PR 14-AUG-2000; 2000US-0225447.

PR 14-AUG-2000; 2000US-0225757.

PR 14-AUG-2000; 2000US-0225758.

PR 14-AUG-2000; 2000US-0225759.

PR 18-AUG-2000; 2000US-0226279.

PR 22-AUG-2000; 2000US-0226681.

PR 22-AUG-2000; 2000US-0226686.

PR 23-AUG-2000; 2000US-0227182.

PR 23-AUG-2000; 2000US-0227009.

PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.

PR 01-SEP-2000; 2000US-0229343.

PR 01-SEP-2000; 2000US-0229344.

PR 01-SEP-2000; 2000US-0229345.

PR 05-SEP-2000; 2000US-0229509.

PR 05-SEP-2000; 2000US-0229513.

PR 06-SEP-2000; 2000US-0230437.

PR 08-SEP-2000; 2000US-0230438.

PR 08-SEP-2000; 2000US-0231242.

PR 08-SEP-2000; 2000US-0231243.

PR 08-SEP-2000; 2000US-0231244.

PR 08-SEP-2000; 2000US-0231413.

PR 08-SEP-2000; 2000US-0231414.

PR 08-SEP-2000; 2000US-0232080.

PR 12-SEP-2000; 2000US-0232081.

PR 14-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.

PR 14-SEP-2000; 2000US-0232398.

PR 14-SEP-2000; 2000US-0232399.

PR 14-SEP-2000; 2000US-0232400.

PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.

PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.

PR 21-SEP-2000; 2000US-0234223.

PR 21-SEP-2000; 2000US-0234274.

PR 25-SEP-2000; 2000US-0234997.

PR 25-SEP-2000; 2000US-0234998.

PR 26-SEP-2000; 2000US-0235484.

PR 27-SEP-2000; 2000US-0235634.

PR 27-SEP-2000; 2000US-0235636.

Matches	334;	Conservative	0;	Mismatches	7;	Indels	0;	Gaps	0;
Qy	1366	GATAAGCCCTTAGCACCAGCTTAGACACCTTCCAAGAACGAGCCCGCTGATGCAAGAT							1425
Db	372	GATAAGCCCTTAGCACCAGCTTAGACACCTTCCAAGAACGAGCCCGCTGATGCAAGAT							431
Qy	1426	GGCAGATCTGATACCCATTAGAGCCCGGAGAAATTCCTTCTGGATCCCGAGTTTGCAGCA							1485
Db	432	GGCAGATCTGATACCCATTAGAGCCCGGAGAAATTCCTTCTGGATCCCGAGTTTGCAGCA							491
Qy	1486	AACCCACACCCAGCTCACACAGCAAAACAATGGACAGGCCCGAGGCTGAAGCAAAC							1545
Db	492	AACCCACACCCAGCTCACACAGCAAAACAATGGACAGGCCCGAGGCTGAAGCAAAC							551
Qy	1546	AGTGCCCTTCTGGCTGTGTGGAGCCTCCCGAGTAACCACTATTATTTTACCTCTTT							1605
Db	552	AGTGCCCTTCTGGCTGTGTGGAGCCTCCCGAGTAACCACTATTATTTTACCTCTTT							611
Qy	1606	CCCAAACTGGAGCATTTATGCCTAGGCTGTGTCAAGAAATCTGTTAGTCCTCTCCTTCT							1665
Db	612	CCCAAACTGGAGCATTTATGCCTAGGCTGTGTCAAGAAATCTGTTAGTCCTCTCCTTCT							671
Qy	1666	CAATAAAGCATCTTCAAGCTTGTAAAAAATAAANGA							1706
Db	672	CAATAAAGCATCTTCAAGCTTAAAAAATAAANGA							712
RESULT 15									
AAS27685/C									
ID	AAS27685 standard; DNA; 2643 BP.								
XX	AC AAS27685;								
XX	DT 07-NOV-2001 (first entry)								
XX	DE DNA encoding novel signal transduction pathway protein, Seq ID 1345.								
XX	Neuroprotective; cytostatic; dermatological; immunosuppressive; tumour; antiinflammatory; anti-HIV; antibacterial; antiinflammatory; cancer; immune system disorder; rheumatoid arthritis; inflammatory condition; organ transplant rejection; infection; hepatitis C; blood disorder; sickle cell anaemia; hyperproliferative disorder; Gaucher's disease; neurodegenerative disorder; Alzheimer's disease; Parkinson's disease; chromosomal abnormality; Down syndrome; ischaemia; renal disorder; cardiovascular; respiratory; wound healing; endocrine; Addison's disease; reproductive system; gastrointestinal; liver disorder; AIDS; ds; acquired immune deficiency syndrome.								
OS	Homo sapiens.								
XX	WO200154733-A1.								
PN	02-AUG-2001.								
XX	17-JAN-2001; 2001WO-US01312.								
PF	31-JAN-2000; 2000US-0179065.								
PR	04-FEB-2000; 2000US-0180628.								
PR	24-FEB-2000; 2000US-0184664.								
PR	02-MAR-2000; 2000US-0186350.								
PR	16-MAR-2000; 2000US-0189874.								
PR	17-MAR-2000; 2000US-0190076.								
PR	18-APR-2000; 2000US-0198123.								
PR	19-MAY-2000; 2000US-0205515.								
PR	07-JUN-2000; 2000US-0209467.								
PR	28-JUN-2000; 2000US-0214886.								
PR	30-JUN-2000; 2000US-0215135.								
PR	07-JUL-2000; 2000US-0216647.								
PR	07-JUL-2000; 2000US-0216880.								
PR	11-JUL-2000; 2000US-0217487.								
PR	11-JUL-2000; 2000US-0217496.								
PR	14-JUL-2000; 2000US-0218290.								
PR	26-JUL-2000; 2000US-0220963.								

PR	26-JUL-2000;	2000US-0220964.
PR	14-AUG-2000;	2000US-0224518.
PR	14-AUG-2000;	2000US-0224519.
PR	14-AUG-2000;	2000US-0225213.
PR	14-AUG-2000;	2000US-0225214.
PR	14-AUG-2000;	2000US-0225266.
PR	14-AUG-2000;	2000US-0225267.
PR	14-AUG-2000;	2000US-0225268.
PR	14-AUG-2000;	2000US-0225270.
PR	14-AUG-2000;	2000US-0225447.
PR	14-AUG-2000;	2000US-0225757.
PR	14-AUG-2000;	2000US-0225758.
PR	14-AUG-2000;	2000US-0225759.
PR	18-AUG-2000;	2000US-0226279.
PR	22-AUG-2000;	2000US-0226681.
PR	22-AUG-2000;	2000US-0226868.
PR	22-AUG-2000;	2000US-0227182.
PR	23-AUG-2000;	2000US-0227009.
PR	30-AUG-2000;	2000US-0228924.
PR	01-SEP-2000;	2000US-0229287.
PR	01-SEP-2000;	2000US-0229343.
PR	01-SEP-2000;	2000US-0229344.
PR	01-SEP-2000;	2000US-0229345.
PR	05-SEP-2000;	2000US-0229509.
PR	05-SEP-2000;	2000US-0229513.
PR	06-SEP-2000;	2000US-0230437.
PR	06-SEP-2000;	2000US-0230438.
PR	08-SEP-2000;	2000US-0231242.
PR	08-SEP-2000;	2000US-0231243.
PR	08-SEP-2000;	2000US-0231244.
PR	08-SEP-2000;	2000US-0231413.
PR	08-SEP-2000;	2000US-0231414.
PR	08-SEP-2000;	2000US-0232080.
PR	08-SEP-2000;	2000US-0232081.
PR	12-SEP-2000;	2000US-0231968.
PR	14-SEP-2000;	2000US-0232397.
PR	14-SEP-2000;	2000US-0232398.
PR	14-SEP-2000;	2000US-0232399.
PR	14-SEP-2000;	2000US-0232400.
PR	14-SEP-2000;	2000US-0232401.
PR	14-SEP-2000;	2000US-0233063.
PR	14-SEP-2000;	2000US-0233064.
PR	21-SEP-2000;	2000US-0233065.
PR	21-SEP-2000;	2000US-0234223.
PR	21-SEP-2000;	2000US-0234274.
PR	25-SEP-2000;	2000US-0234997.
PR	25-SEP-2000;	2000US-0234998.
PR	26-SEP-2000;	2000US-0235484.
PR	27-SEP-2000;	2000US-0235834.
PR	27-SEP-2000;	2000US-0235836.
PR	29-SEP-2000;	2000US-0236327.
PR	29-SEP-2000;	2000US-0236367.
PR	29-SEP-2000;	2000US-0236368.
PR	29-SEP-2000;	2000US-0236369.
PR	02-OCT-2000;	2000US-0236370.
PR	02-OCT-2000;	2000US-0236802.
PR	02-OCT-2000;	2000US-0237037.
PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	02-OCT-2000;	2000US-0237040.
PR	13-OCT-2000;	2000US-0239935.
PR	13-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	20-OCT-2000;	2000US-0241826.
PR	01-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.

PR 08-NOV-2000; 2000US-0246476.
PR 08-NOV-2000; 2000US-0246477.
PR 08-NOV-2000; 2000US-0246478.
PR 08-NOV-2000; 2000US-0246523.
PR 08-NOV-2000; 2000US-0246524.
PR 08-NOV-2000; 2000US-0246525.
PR 08-NOV-2000; 2000US-0246526.
PR 08-NOV-2000; 2000US-0246527.
PR 08-NOV-2000; 2000US-0246528.
PR 08-NOV-2000; 2000US-0246532.
PR 08-NOV-2000; 2000US-0246609.
PR 08-NOV-2000; 2000US-0246610.
PR 08-NOV-2000; 2000US-0246611.
PR 08-NOV-2000; 2000US-0246613.
PR 17-NOV-2000; 2000US-0249207.
PR 17-NOV-2000; 2000US-0249208.
PR 17-NOV-2000; 2000US-0249209.
PR 17-NOV-2000; 2000US-0249210.
PR 17-NOV-2000; 2000US-0249211.
PR 17-NOV-2000; 2000US-0249212.
PR 17-NOV-2000; 2000US-0249213.
PR 17-NOV-2000; 2000US-0249214.
PR 17-NOV-2000; 2000US-0249215.
PR 17-NOV-2000; 2000US-0249216.
PR 17-NOV-2000; 2000US-0249217.
PR 17-NOV-2000; 2000US-0249218.
PR 17-NOV-2000; 2000US-0249244.
PR 17-NOV-2000; 2000US-0249245.
PR 17-NOV-2000; 2000US-0249264.
PR 17-NOV-2000; 2000US-0249265.
PR 17-NOV-2000; 2000US-0249297.
PR 17-NOV-2000; 2000US-0249299.
PR 01-DEC-2000; 2000US-0250160.
PR 01-DEC-2000; 2000US-0250391.
PR 05-DEC-2000; 2000US-0251030.
PR 05-DEC-2000; 2000US-0251988.
PR 06-DEC-2000; 2000US-0256719.
PR 06-DEC-2000; 2000US-0251479.
PR 08-DEC-2000; 2000US-0251856.
PR 08-DEC-2000; 2000US-0251868.
PR 08-DEC-2000; 2000US-0251869.
PR 08-DEC-2000; 2000US-0251989.
PR 11-DEC-2000; 2000US-0251990.
PR 05-JAN-2001; 2001US-0259678.
XX (HUMA-) HUMAN GENOME SCI INC.
PA
XX
PI Rosen CA, Barash SC, Ruben SM;
DR WPI; 2001-465460/50.
XX
PT Novel polypeptides useful for diagnosing, treating, preventing and/or
PT prognosing disorders related to the proteins, including cancers, immune
PT disorders and neuronal disorders
XX
PS Claim 1; SEQ ID No 1345; 880pp; English.
XX

CC The invention relates to novel isolated polypeptides (I), and
CC polynucleotides (II). (i), (ii) and the antibody to (i) are useful for
CC diagnosing, preventing and treating diseases including immune system
CC disorders (e.g. congenital and acquired immunodeficiencies, autoimmune
CC disorders (e.g. rheumatoid arthritis), inflammatory conditions, organ
CC transplant rejections and graft versus host disease, infectious diseases
CC (e.g. hepatitis C), bleeding disorders, haemoglobin abnormalities and
CC other blood-related disorders (sickle cell anaemia), myeloproliferative
CC disorders, primary haematopoietic disorders, hyperproliferative
CC disorders (e.g. Gaucher's disease and cancer), neurodegenerative
CC disorders (e.g. Alzheimer's disease, Parkinson's disease), chromosomal
CC abnormalities (Down syndrome), ischaemic injury (e.g. stroke), renal
CC disorders (e.g. glomerulonephritis), cardiovascular disorders
CC (e.g. arrhythmia), respiratory disorders, dermatological disorders, in

CC wound healing, epithelial cell proliferation, endocrine disorders (e.g.
CC Addison's disease), reproductive system disorders, gastrointestinal
CC disorder (inflammatory disorders), liver disorders (cirrhosis),
CC as stimulators of B-cell responsiveness to pathogens, activators of
CC T-cells, to induce higher affinity antibodies, and as a means to induce
CC tumour proliferation in pathologies e.g. acquired immune deficiency
CC syndrome (AIDS). AAS26976-AAS27850 represent novel signal transduction
CC pathway protein coding sequences and PCR primers of the invention.
XX
SQ Sequence 2643 BP; 562 A; 729 C; 705 G; 647 T; 0 other;

Query Match 18.9%; Score 324; DB 22; Length 2643;
Best Local Similarity 100.0%; Pred. No. 4.8e-83;
Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1366 GATAAGCCCTTAGCCAGCTTAGACACCTCCCAAGAACCCAGCCCGCTGATGCAAGAT 1425
Db 324 GATAAGCCCTTAGCCAGCTTAGACACCTCCCAAGAACCCAGCCCGCTGATGCAAGAT 265
Qy 1426 GGCAGATCTGATACCCCATTTAGAGCCCGGAGAAATTCCTCTTCTGGATCCAGTTTGCAGCA 1485
Db 264 GGCAGATCTGATACCCCATTTAGAGCCCGGAGAAATTCCTCTTCTGGATCCAGTTTGCAGCA 205
Qy 1486 AACCCACACCCAGCTCAGACAGCAAAACAATGGACAGGCCAGAGGCTGAAGCAAAAC 1545
Db 204 AACCCACACCCAGCTCAGACAGCAAAACAATGGACAGGCCAGAGGCTGAAGCAAAAC 145
Qy 1546 AGTGTCCTTCTGGCTGTGTGGAGCCCTCCCGAGTAACCACTATTTATTTTACTCTTT 1605
Db 144 AGTGTCCTTCTGGCTGTGTGGAGCCCTCCCGAGTAACCACTATTTATTTTACTCTTT 85
Qy 1606 CCCAAACCTGGAGCATTTATGCTAGGCTTGTCAAGAAATCTGTTCAAGTCCCTCTCTTCT 1665
Db 84 CCCAAACCTGGAGCATTTATGCTAGGCTTGTCAAGAAATCTGTTCAAGTCCCTCTCTTCT 25
Qy 1666 CAATAAAAGCATCTTCAAGCTTGT 1689
Db 24 CAATAAAAGCATCTTCAAGCTTGT 1

Search completed: February 25, 2003, 07:43:06
Job time : 284 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 07:37:12 ; Search time 60 Seconds
(without alignments)
8740.288 Million cell updates/sec

Title: US-09-762-021a-1
Perfect score: 1710
Sequence: 1 ggcagagcgactgaagacca.....aaaaaaataangataaa 1710

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 1533381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NH:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	975.8	57.1	1651	4	US-09-484-970B-48
2	155.6	9.1	197	1	US-08-700-575-40
3	93.2	5.5	1635	4	US-09-484-970B-51
4	49.8	2.9	3245	1	US-07-935-311A-3
5	49.8	2.9	3245	1	US-08-368-079-3
6	49.8	2.9	3245	5	PCT-US93-07993-3
7	47.4	2.8	1457	4	US-09-444-053-3
8	44.4	2.6	1803	4	US-09-006-428A-18
9	44.4	2.6	1858	4	US-09-006-428A-16
10	43.8	2.6	7218	1	US-08-232-463-14
11	39.6	2.3	70000	4	US-09-851-896-3
12	35.6	2.1	3240	4	US-09-851-896-10
13	35.6	2.1	5661	4	US-08-938-105-2
14	35.4	2.1	6530	2	US-08-146-930-1
15	35.4	2.1	6530	3	US-08-458-240-1
16	35.4	2.1	6530	5	PCT-US93-03993-1
17	35.4	2.1	15894	1	US-08-348-891A-1
18	35.4	2.1	15894	1	US-08-905-817-1
19	35.2	2.1	16442	3	US-08-781-891-208
20	34.4	2.0	4403765	4	US-09-103-840A-2
21	34.4	2.0	4411529	4	US-09-103-840A-1
22	34.2	2.0	480	5	PCT-US95-08295-14
23	34.2	2.0	7218	1	US-08-232-463-14
24	34	2.0	241	3	US-08-866-340-17
25	34	2.0	4411529	4	US-09-103-840A-1
26	33.8	2.0	1296	1	US-07-816-283-9
27	33.8	2.0	1296	1	US-08-417-103-9

RESULT 1

US-09-484-970B-48
; Sequence 48, Application US/09484970B
; Patent No. 6426186

; GENERAL INFORMATION:
; APPLICANT: Jones, Karen A.
; APPLICANT: Volkmut, Wayne
; APPLICANT: Walker, Michael G.
; TITLE OF INVENTION: BONE REMODELING GENES
; FILE REFERENCE: PB-0014 US
; CURRENT APPLICATION NUMBER: US/09/484,970B
; CURRENT FILING DATE: 2000-01-18
; NUMBER OF SEQ ID NOS: 172
; SOFTWARE: PERL Program
; SEQ ID NO 48

; LENGTH: 1651
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Inceyte ID No. 6426186 337950.5CB1

; NAME/KEY: unsure
; LOCATION: 1531, 1570, 1647
; OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-48

Query Match 57.1%; Score 975.8; DB 4; Length 1651;

Best Local Similarity 91.2%; Pred. No. 2.6e+266;

Matches 1106; Conservative 0; Mismatches 14; Indels 93; Gaps 3;

QY 1 GGCAGCGCGACTGAAGCAGCCCTGCAGAGGCTCTCGAGAGAGAGCTGGAGCAAGACC 60

Db 520 GGCAGCGCGACTGAAGCAGCCCTGCAGAGGCTCTCGAGAGAGAGCTGGAGCAAGACC 579

QY 61 TCGACTTGGAGGCTTCAGCCAGCCAGGACATGAGGGGGCTCTGATGGAAGGCC 120

Db 580 TCGACTTGGAGGCTTCAGCCAGCCAGGACATGAGGGGGCTCTGATGGAAGGCC 639

QY 121 GCTCCCTATGGAGCAGCAGCTATCTGGAGCGGGATCCCTCCAGAACAGCCCAACCA 180

Db 640 GCTCCCTATGGAGCAGCAGCTATCTGGAGCGGGATCCCTCCAGAACAGCCCAACCA 699

QY 181 GAGGACCTTAGAGCAGCAGCTCCACCATCCCAAGGCCCTGCGACGCCACACAGTGC 240

Db 700 GAGGACCTTAGAGCAGCAGCTCCACCATCCCAAGGCCCTGCGACGCCACACAGTGC 759

QY 241 CCGAGAACCAAGTGCCTTTACTCTGCTCTCTCCAAAGCGGTCTCTTCCCGGAGGACC 300

Db 760 CCGAGAACCAAGTGCCTTTACTCTGCTCTCTCCAAAGCGGTCTCTTCCCGGAGGACC 819

ALIGNMENTS

SEQ ID NO 51
LENGTH: 1635
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
OTHER INFORMATION: Incyte ID No. 6426186 403869.2CBI
LOCATION: 884, 1049, 1069, 1629, 1632
OTHER INFORMATION: a, t, c, g, or other
US-09-484-970B-51

Query Match 5.5%; Score 93.2; DB 4; Length 1635;
Best Local Similarity 49.1%; Pred. No. 1.4e-16;
Matches 336; Conservative 0; Mismatches 338; Indels 10; Gaps 3;

QY 431 CTCACCCAGGCACATGCTGCTCCAGAGATCAAGTACAGCTTCAACCTCCTG 490
DB 65 CCCTCGGAGCGGAGTACCGAGTCTGCAGAGATCAAGTACGCTTTCAGCCTG 124
QY 491 GGAAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCTGAGCTCGTACACATCCTCTTC 550
DB 125 GCCCGCTGGCGGCAACATCGCGACCCCTCTCTCCGAGCTGTTGCACTTCTTTTC 184
QY 551 AAGTCCCTGAATTCATCTCGTGGCAGTGGCCCTGAGCTGGCTAGCAGCCCAAGTGATC 610
DB 185 GGGCTCTGCAGATGATGTGAACACCTCGGGGGCGGAGTTCGGGAGCAGGTGTCGG 244
QY 611 TCACCCCTCTCACCCCTAAAGTATCAACTGCTACAGTCTGCTTAAGCCACCTGAG 670
DB 245 CGGCGCATCTGACATCGGATGGCGTGGCGCTGCTGGGACAACTCACTTCCACGTGAA 304
QY 671 AGTAACCTTTGGAGGGTGGCGCCAGCCTGGACCACTAGCCGGCGCGTGGACAGGC 730
DB 305 AACAGCTCTGGACCTCGCTGGGGGACTCGTGGACCCCGCGGCTGGAGCTGTCCCG 364
QY 731 GATGAGCCCTGGCCCTACCAACCCACATCTTCAGATGACTGGCAACTTCCAGAGCCCTCC 790
DB 365 GAGGAGGACCCCATACAGACCCGAGTCTTTCAGCGGCTGGGAGCGCGGTCACTGAC 424
QY 791 AGCAACACCCCTTAGATACCAAGGACCTGTTTCCCTTGGCGGGGAAAGTATAGTTA 850
DB 425 CCGCAGAGCGCG---GCCTGGGAGGACCCAGTTGAGAAACAGC---TACAGCAGCAGCGG 478
QY 851 GGGAGCACCTCACACTTCTCAGGAGGAGACACACCAACATGACCTAGCCTGGGGAC 910
DB 479 AGGCGCGGAGCAAAAGCGCCCGAGGTGCTGCTCAATGGTACCAGACACTTGGAGC 538
QY 911 CCCAACTCCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGAAAATGCAAGTCTTG 970
DB 539 CAGAACTGAGCCCTCAGCTG---GAGTCAGAGACAGCAAGAAATGGGCTCTGTGTAAT 594
QY 971 TACAGATTTGAGCTAGAACCCACGCGGAACTGACTGTGTCCAGGAGAGAAAGCTGGAG 1030
DB 595 TATGACTTCCAGGCCCAACAGCAGTGAAGTGTGCTGCAAGCAGCGGGAGCTACTGGAG 654
QY 1031 GTTCTGGACACACCAAGCGGTGTGCTGGTGAAGAAATGAGCGGACGAGCGGCTAC 1090
DB 655 GTCTGATGATGATCTGTAAGTGTGAAGGTTCGGACCCAGCGGGCAGGAGGATAT 714
QY 1091 ATTCAGACACATCTCTGGAGCC 1114
DB 715 GTGCCCTACAACATCTCTGACACC 738

RESULT 4
US-07-935-311A-3
Sequence 3, Application US/07935311A
Patent No. 5378809
GENERAL INFORMATION:
APPLICANT: Di Fiore, Pier Paolo
APPLICANT: Fazioli, Francesca
TITLE OF INVENTION: Substrate of the Epidermal Growth

TITLE OF INVENTION: Factor Kinase
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson & Bear
STREET: 620 Newport Center Drive, 16th Floor
CITY: Newport Beach
STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/935,311A
FILING DATE: 19920825
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Israel, Ned A.
REGISTRATION NUMBER: 29,655
REFERENCE/DOCKET NUMBER: NIH035.001A
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619) 235-8550
TELEFAX: (619) 235-0176
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 3245 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 246..2708
US-07-935-311A-3

Query Match 2.9%; Score 49.8; DB 1; Length 3245;
Best Local Similarity 47.8%; Pred. No. 0.00037;
Matches 144; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 444 AGTACATTTGACTGCTTCCAGAGATCAAGTACAGCTTCAACCTCTCGGAGGCTGGCCA 503
DB 1225 AGTTGTTGACTGTTTCCAGAAAGTTTAAACATGGATTCAACCTTCTGGCCAAGTTGAAGT 1284
QY 504 CTGGCTGAGGAGACAAAGTCCCTGAGCTCGTACACATCTTCTCAAGTCCCTGAAC 563
DB 1285 CCATATCCAGAACCCGAGTCTCAGATCTGGTTCATTTTGTCTTACTCCACTAAATA 1344
QY 564 TCATCTCTGGCCAGGTGCTGAGGCTGGCTTAGCAGCCCAAGTGATCTCACCCCTCCTCA 623
DB 1345 TGGTGTCCAGGCAACAGTGGCCCTGAAGTGGCCAGTTCGGTACTCAGCCACTGTTGA 1404
QY 624 CCCTAAAGCTATCAACCTGCTACAGTCTCTTAAGCCCACTGAGAGTAACCTTTGGA 683
DB 1405 CAAAAGACACAGTTGATTTCTTAAACTACACAGCCACTGCGGAGGAAACGGAAGTGTGA 1464
QY 684 TGGGTTGGCCCAACCTGACCACTAGCCGGCCGAGTGGACAGGCGATGAGCCCTGC 743
DB 1465 TGCTACTGGGAGATAGTTGGTGAAAGTGAGAGCAGAGTGGCCGAAAGAACAGTTCATCC 1524
QY 744 C 744
DB 1525 C 1525

RESULT 5
US-08-368-079-3
Sequence 3, Application US/08368079
Patent No. 5610018
GENERAL INFORMATION:
APPLICANT: Di Fiore, Pier Paolo
APPLICANT: Fazioli, Francesca

```

: TITLE OF INVENTION: eps8, A Substrate for the Epidermal Growth Factor
:
: TITLE OF INVENTION: Kinase, Antibodies Thereeto, and Methods of Use Thereof
:
: NUMBER OF SEQUENCES: 6
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Knobbe, Martens, Olson & Bear
:
: STREET: 620 Newport Center Drive, 16th Floor
:
: CITY: Newport Beach
:
: STATE: CA
:
: COUNTRY: USA
:
: ZIP: 92660
:
: COMPUTER READABLE FORM:
:
: MEDIUM TYPE: Diskette
:
: COMPUTER: IBM Compatible
:
: OPERATING SYSTEM: DOS
:
: SOFTWARE: FastSeq Version 1.5
:
: CURRENT APPLICATION DATA:
:
: APPLICATION NUMBER: US/08/368,079
:
: FILING DATE: 03-JAN-1995
:
: CLASSIFICATION: 530
:
: PRIOR APPLICATION DATA:
:
: APPLICATION NUMBER: 07/935,311
:
: FILING DATE: 25-AUG-1992
:
: ATTORNEY/AGENT INFORMATION:
:
: NAME: Israelson, Ned A.
:
: REGISTRATION NUMBER: 29,655
:
: REFERENCE/DOCKET NUMBER: NIH035.001DV1
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: (619) 235-8550
:
: TELEFAX: (619) 235-0176
:
: INFORMATION FOR SEQ ID NO: 3:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 3245 base pairs
:
: TYPE: nucleic acid
:
: STRANDEDNESS: double
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
:
: FEATURE:
:
: NAME/KEY: CDS
:
: LOCATION: 246..2708
:
: US-08-368-079-3

```

```

: GENERAL INFORMATION:
: APPLICANT: The Government of the United States, as represented by the
: APPLICANT: Secretary of Health and Human Services
: TITLE OF INVENTION: Substrate of the Epidermal Growth Factor Kinase
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Knobbe, Martens, Olson & Bear
: STREET: 620 Newport Center Drive, 16th Floor
: CITY: Newport Beach
: STATE: CA
: COUNTRY: USA
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: PCT/US93/07996
: FILING DATE: 19930825
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (619) 235-8550
: TELEFAX: (619) 235-0176
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3245 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 246..2708
PCT-US93-07996-3

Query Match 2.9%; Score 49.8; DB 5; Length 3245;
Best Local Similarity 47.8%; Pred. No. 0.00037;
Matches 144; Conservative 0; Mismatches 157; Indels 0; Gaps

QY 444 AGTACATTGACTCTTCCAGAGATCAAGTACAGCTTCAACCTCTCTGGGAAGGCTGGCCA 503
Db 1225 AGTTGTGTGACTGTTTCCAGAGATTAAACATGGGATTCAACCTTCTGGCCNAGTTGAAGT 1284

QY 504 CCTGGCTGAAGGAGACAAGTGGCCCTTGAGCTCGTACACATCCTCTTCAAGTCCCTGAACT 563
Db 1285 CCATATCCAGAACCCGAGTGCTTCAGATCTCGTTCAITTTTTTGTTTACTCCCAATAA 1344

QY 564 TCATCTGGCCAGTGCCCTGAGGTGGCGCTAGCAGCCCAAGTATCTCACCCCTCTCTCA 623
Db 1345 TGGTGTGCCAGGCAACAGGTGGCCCTGAACTGGCCAGTTTCGGTACTCAGCCCACTGTGA 1404

QY 624 CCCCTAAAGCTATCAACCTGCTACAGTCTCTGTCCTAAGCCACCTTGAGAGTAACCTTTGGA 683
Db 1405 CAAAGACACAGTTGATTTCTTAAACTACACAGCCACTGCGGAGAGACGGGAAGCTGTGGA 1464

QY 684 TGGGGTTGGGCCCCAGCCTGGACCACTAGCCGGGGCCGACTGACAGAGCGGATGAGCCCTGC 743
Db 1465 TGTCACCTGGGAGATAGTTGGGTCAAGTCAGAGCAGAGTGGCCCGAAGACAGTTTCATCC 1524

QY 744 C 744
Db 1525 C 1525

RESULT 7
US-09-444-053-3
: Sequence 3, Application US/09444053A
: Patent No. 6165728
: GENERAL INFORMATION:
: APPLICANT: Donna T. Ward
: APPLICANT: Lex M. Cowser
: TITLE OF INVENTION: ANTISENSE MODULATION OF NCK-2 EXPRESSION
: FILE REFERENCE: RTS-0122

```



```
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: pTZgpt-F1s
US-08-232-463-14

Query Match 2.6%; Score 43.8; DB 1; Length 7218;
Best Local Similarity 5.5%; Pred. No. 0.027;
Matches 15; Conservative 152; Mismatches 104; Indels 0; Gaps 0;

Qy 293 GAGGACCAGAGAGGAGGAGGAGTGTGAACCATGCTCAAGGACATGAGCTGTTC 352
Db 1306 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1247
Qy 353 ATGGGAAAGCTGGAGAGCGCCAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 412
Db 1246 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1187
Qy 413 AACAGAGCAGGAGGAGTCTCACCAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 472
Db 1186 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1127
Qy 473 TACAGCTTCAACCTCTGGGAGGCTGGCCAGCTGGCTGAAGGAGAGAGTGCCTGAG 532
Db 1126 RRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRRR 1067
Qy 533 CTGTCACATCTCTTCAAGTCCCTGAAC 563
Db 1066 ATCGCAAGCTCCCTCGACCTGAGCCAAAGCT 1036

RESULT 11
US-09-851-896-3
; Sequence 3, Application US/09851896
; Patent No. 6410325
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEP
; TITLE OF INVENTION: EXPRESSION
; FILE REFERENCE: RTS-0220
; CURRENT APPLICATION NUMBER: US/09/851.896
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 3
; LENGTH: 70000
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
US-09-851-896-3

Query Match 2.3%; Score 39.6; DB 4; Length 70000;
Best Local Similarity 56.0%; Pred. No. 1.3;
Matches 75; Conservative 0; Mismatches 59; Indels 0; Gaps 0;
```

```
Qy 435 CCCAGGCACAGTACATTGACTTGCTTCCAGAAAGATCAAGTACAGCTTCAACCTCCTGGAA 494
Db 68372 CCTGAGCATCTAGGTGACCCCTCTCCCTGGCCGAGACAGATTGAACCCAGCTGG 68431
Qy 495 GGCTGGCCACCTGGCTGAAGGAGAGAGAGTGCCTCGTACATCTCTTCTTCAAGT 554
Db 68432 GGACGGACATCATGCTGGATGAGGTGACAGAGTCTGTGTAACGCCCTCTGGGAGA 68491
Qy 555 CCCTGAACCTTCATC 568
Db 68492 CCGAGGTCTACATC 68505

RESULT 12
US-09-851-896-10
; Sequence 10, Application US/09851896
; Patent No. 6410325
; GENERAL INFORMATION:
; APPLICANT: C. Frank Bennett
; APPLICANT: Susan M. Freier
; APPLICANT: Andrew T. Watt
; TITLE OF INVENTION: ANTISENSE MODULATION OF PHOSPHOLIPASE A2, GROUP VI (CA2+-INDEP
; TITLE OF INVENTION: EXPRESSION
; FILE REFERENCE: RTS-0220
; CURRENT APPLICATION NUMBER: US/09/851.896
; CURRENT FILING DATE: 2001-05-08
; NUMBER OF SEQ ID NOS: 89
; SEQ ID NO 10
; LENGTH: 3240
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: 2770
; OTHER INFORMATION: unknown
; NAME/KEY: CDS
; LOCATION: (137)....(2557)
US-09-851-896-10

Query Match 2.1%; Score 35.6; DB 4; Length 3240;
Best Local Similarity 60.2%; Pred. No. 3.8;
Matches 59; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

Qy 471 AGTACAGCTTCAACCTCTCTGGGAGGCTGGCCACCTGGCTGAAGGAGAGACAGTGCCTCG 530
Db 2406 ACTTCAGATTGAACCCCGAGCTGGGAGGAGACATCATGCTGGATGAGGTGAGTGCACACAG 2465
Qy 531 AGCTCGTACATCTCTTCAAGTCCCTGAACCTTCATC 568
Db 2466 TGCTGGTCAACGCCCTCTGGGAGACCCGAGGTCTACATC 2503

RESULT 13
US-08-938-105-2
; Sequence 2, Application US/08938105
; Patent No. 6353151
; GENERAL INFORMATION:
; APPLICANT: Leinwand, Leslie A.
; APPLICANT: Vikstrom, Karen L.
; TITLE OF INVENTION: TRANSGENIC MODEL FOR HEART FAILURE
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sheridan Ross P.C.
; STREET: 1700 Lincoln St., Suite 3500
; CITY: Denver
; STATE: CO
; COUNTRY: U.S.A.
; ZIP: 80203
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```



```

; APPLICATION NUMBER: US/08/458,240
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/146,930
; FILING DATE:
; APPLICATION NUMBER: 07/876,286
; FILING DATE: April 30, 1992
; APPLICATION NUMBER: No. 6143727 yet assigned (204/132)
; FILING DATE: October 29, 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 204/152
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6530 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-458-240-1

Query Match      2.1%; Score 35.4; DB 3; Length 6530;
Best Local Similarity 49.7%; Pred. No. 6.2;
Matches 90; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 47 CTGAGACAAAGACCTCGACTTGGAGGCCCTTCAGCCAGCGCCAGACAGATGGAGGGGGCCT 106
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3478 CCGGAGGAAACCACTCCCGAGCTCGTGGCGCGCCACCGGAATAGCCGCCGCCGCCGCG 3419

QY 107 GCTATGGAAGCCCTCCCTATGAGGAGGACGCGTATCTGGAGCGCGGGATCCCTCCA 166
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3418 CAGCTGGATCCACGCCCACTGAGTAGCGCGCCACAGCTAGAGCGCGCGCTGCTTCCA 3359

QY 167 GAACAGCCCCACCCAGGAGGACCGCTAGAGCACAGCCCTCCCAAGCCCTGCGCA 226
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 3358 CCGCGCGCGGAATAGCCCGCTCCGCGAGCTGGAGCCACCAGCCGCCACAGAGTACTTCCG 3299

QY 227 C 227
   |
DB 3298 C 3298
```

Search completed: February 25, 2003, 16:13:39
Job time : 226 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 08:44:57 ; Search time 81 seconds

(without alignments)

11856.832 Million cell updates/sec

Title: US-09-762-021a-1

Perfect score: 1710

Sequence: 1 ggcagagcgactgaagacca.....aaaaaaaaataangataaa 1710

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 442118 seqs, 280819700 residues

Total number of hits satisfying chosen parameters: 884236

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1508.6	88.2	2148	10	US-09-998-598-2588
2	829.6	48.5	872	9	US-09-764-868-144
3	673.8	39.4	1265	9	US-09-764-868-143
4	487	28.5	499	10	US-09-998-598-188
5	411.2	24.0	447	10	US-09-998-598-2183
6	389	22.7	401	10	US-09-880-107-290
7	335.4	19.6	572	9	US-09-764-868-522
8	324	18.9	2643	9	US-09-764-868-1345
9	119.2	7.0	216	10	US-09-925-299-321
10	68.8	4.0	595	10	US-09-925-299-578
11	59	3.5	245	10	US-09-998-598-2415
12	56.4	3.3	1639	9	US-09-764-868-99
13	56.4	3.3	1661	9	US-09-764-868-512
14	55	3.2	514	10	US-09-833-790-43
15	44.4	2.5	488	10	US-09-867-701-10401
16	39.2	2.3	1466	10	US-09-925-300-531
17	39.2	2.3	23580	10	US-09-764-860-990
18	38.2	2.2	449	10	US-09-967-768A-23
19	38	2.2	1110	10	US-09-893-737-49

20	36.8	2.2	1329	9	US-10-011-588-22	Sequence 22, Appl
21	36.8	2.2	2559	9	US-10-011-588-24	Sequence 24, Appl
22	36.6	2.1	2928	9	US-09-764-868-142	Sequence 142, App
23	36.4	2.1	252	10	US-09-969-708-484	Sequence 484, App
24	36.4	2.1	252	10	US-09-880-107-949	Sequence 949, App
25	36.4	2.1	1216	10	US-09-917-800A-1453	Sequence 1453, App
26	36.2	2.1	359	9	US-09-764-868-521	Sequence 521, App
27	36	2.1	355	10	US-09-964-824A-518	Sequence 518, App
28	36	2.1	370	10	US-09-969-708-456	Sequence 456, App
29	35.8	2.1	449	10	US-09-764-847-328	Sequence 328, App
30	35.4	2.1	681	12	US-10-002-600-19	Sequence 19, Appl
31	35.4	2.1	1042	12	US-10-002-600-20	Sequence 20, Appl
32	35.4	2.1	1109	12	US-10-002-600-21	Sequence 21, Appl
33	35.4	2.1	1123	9	US-09-822-846-357	Sequence 357, App
34	35.4	2.1	4398	9	US-10-105-989-1	Sequence 1, Appl
35	35.2	2.1	108	10	US-09-919-580-710	Sequence 710, App
36	35.2	2.1	2070	9	US-09-932-367A-27	Sequence 27, Appl
37	35.2	2.1	2160	9	US-09-932-367A-9	Sequence 9, Appl
38	35.2	2.1	2190	9	US-09-932-367A-11	Sequence 11, Appl
39	35.2	2.1	2515	9	US-09-932-367A-25	Sequence 25, Appl
40	35.2	2.1	2540	9	US-09-932-367A-26	Sequence 26, Appl
41	35.2	2.1	8667	9	US-09-932-367A-22	Sequence 22, Appl
42	35	2.0	321	10	US-09-864-761-25530	Sequence 25530, A
43	35	2.0	575	10	US-09-864-761-8864	Sequence 8864, App
44	35	2.0	2451	10	US-09-771-161A-4	Sequence 4, Appl
45	35	2.0	3684	10	US-09-822-849A-76	Sequence 76, Appl

ALIGNMENTS

RESULT 1

US-09-998-598-2588

; Sequence 2588, Application US/09998598

; Patent No. US20020150922A1

; GENERAL INFORMATION:

; APPLICANT: Stolk, John A.

; APPLICANT: Xu, Jiangchun

; APPLICANT: Chenault, Ruth A.

; APPLICANT: Meagher, Madelein Joy

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER

; FILE REFERENCE: 210121.561

; CURRENT APPLICATION NUMBER: US/09/998, 598

; CURRENT FILING DATE: 2001-11-16

; NUMBER OF SEQ ID NOS: 2606

; SOFTWARE: Corixa Invention Disclosure Database

; SEQ ID NO 2588

; LENGTH: 2148

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-998-598-2588

Query Match 88.2%; Score 1508.6; DB 10; Length 2148;

Best Local Similarity 94.4%; Pred. No. 0;

Matches 1611; Conservative 0; Mismatches 5; Indels 90; Gaps 1;

Qy 1 GCGAGCGGCTGAAGACCCAGCCCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGACC 60

Db 532 GCGAGCGGCTGAAGACCCAGCCCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGACC 591

Qy 61 TCGACTTGGAGGCTTCAGCCAGGACAGATGAGGGGGCTCTATGGAAGGCC 120

Db 592 TCGACTTGGAGGCTTCAGCCAGGACAGATGAGGGGGCTCTATGGAAGGCC 651

Qy 121 GCTCCCTATGGAGCAGCAGCTATCTGGAGCGGGGATCCCTCCAGAACAGCCCCACCA 180

Db 652 GCTCCCTATGGAGCAGCAGCAGCTATCTGGAGCGGGGATCCCTCCAGAACAGCCCCACCA 711

Qy 181 GAGGACCTTAGACACAGCTCCACCATCCCCAAGCCCTGCGCCAGCCACACAGTGC 240

Db 712 GAGGACCTTAGACACAGCTCCACCATCCCCAAGCCCTGCGCCAGCCACACAGTGC 771


```
QY 1046 AAGCGGTGGTGGCTGATGAAGATGAGCGGGACGGAGCGGCTACATTTCAAGCAATC 1105
Db 210 AAGCGGTGGTGGCTGATGAAGATGAGCGGGACGGAGCGGCTACATTTCAAGCAATC 269
QY 1106 CTGAGGCCCTTACAGCGCGGGACCCCTGGGACCCAGGCGCAGTCACCTCTCGGGTTCCA 1165
Db 270 CTGAGGCCCTTACAGCGCGGGACCCCTGGGACCCAGGCGCAGTCACCTCTCGGGTTCCA 329
QY 1166 ATGCTTCGACTTAGCTCGAGGSCCTGAAGAGGTCAAGAGTGGCTGCGAGGAGAGAACTTC 1225
Db 330 ATGCTTCGACTTAGCTCGAGGSCCTGAAGAGGTCAAGAGTGGCTGCGAGGAGAGAACTTC 389
QY 1226 TCCACTGCCACGGTAGGACACTTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAGA 1285
Db 390 TCCACTGCCACGGTAGGACACTTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAGA 449
QY 1286 CCTGGGAGCTTACAGATGCTATGTCACAGAGGCCCCAGCAATCTCTCCGGCTGGAG 1345
Db 450 CCTGGGAGCTTACAGATGCTATGTCACAGAGGCCCCAGCAATCTCTCCGGCTGGAG 509
QY 1346 GCTGTCAAGAGGATGCTGGGATGAAGCCCTTAGGCACACAGCTTAGACACCTTCCAAGAACC 1405
Db 510 GCTGTCAAGAGGATGCTGGGATGAAGCCCTTAGGCACACAGCTTAGACACCTTCCAAGAACC 569
QY 1406 AGGCCCCGCTGATGCAAGATGGCAGATCTGATACCCATTAGAGCCCCGAGAAATTCCTTT 1465
Db 570 AGGCCCCGCTGATGCAAGATGGCAGATCTGATACCCATTAGAGCCCCGAGAAATTCCTTT 629
QY 1466 CTGATCCCACTTTGACAGAAACCCACACCCAGCTCACACAGCAAAAACATGGACAG 1525
Db 630 CTGATCCCACTTTGACAGAAACCCACACCCAGCTCACACAGCAAAAACATGGACAG 689
QY 1526 GCCCAGAGGCTGAAGCAACAGTCCCTTCTGGCTGTGTTGGAGCCTCCCCAGTAACCA 1585
Db 690 GCCCAGAGGCTGAAGCAACAGTCCCTTCTGGCTGTGTTGGAGCCTCCCCAGTAACCA 749
QY 1586 CCTATTATTTTACCTCTTTCCCAAACTGGAGCATTTATGCCCTAGGCTTTGCAAGATC 1645
Db 750 CCTATTATTTTACCTCTTTCCCAAACTGGAGCATTTATGCCCTAGGCTTTGCAAGATC 809
QY 1646 TGTTCAGTCCCTCTCCCTCTCAATAAAGCATCTTCAAGCTTGTAATAAAAAAATA 1702
Db 810 TGTTCAGTCCCTCTCCCTCTCAATAAAGCATCTTCAAGCTTGTAATAAAAAAATA 866
```

RESULT 3

US-09-764-868-143/c

```
; Sequence 143, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 143
; LENGTH: 1265
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (1263)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-868-143
```

Query Match 39.4%; Score 673.8; DB 9; Length 1265;
Best Local Similarity 98.0%; Pred. No. 5e-177;
Matches 678; Conservative 3; Mismatches 11; Indels 0; Gaps 0;

QY 1015 GGGAGAGAAGCTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGTGAAGATGAGGC 1074

```
Db 693 KGSTGACTTGTGCAGGTTCTGACCACAGCAAGCGGTGGTGGTGAAGATGAGGC 634
QY 1075 GGAACGAGCGGCTTACATTTCAAGCAACATCTCGAGCCCTTACAGCCGGGACCCCTGG 1134
Db 633 GGAACGAGCGGCTTACATTTCAAGCAACATCTCGAGCCCTTACAGCCGGGACCCCTGG 574
QY 1135 GACCCAGGGCCAGTCACCCCTCTCGGGTTCCAATGCTTTCGACTTAGCTCGAGGCTGAAGA 1194
Db 573 GACCCAGGGCCAGTCACCCCTCTCGGGTTCCAATGCTTTCGACTTAGCTCGAGGCTGAAGA 514
QY 1195 GGTCAACAGACTGGCTGACGAGCAGAGAATCTTCCACTGCCACGGTGAAGACACTTGGGTC 1254
Db 513 GGTCAACAGACTGGCTGACGAGCAGAGAATCTTCCACTGCCACGGTGAAGACACTTGGGTC 454
QY 1255 CCTGACGGGAGCCAGCTACTTTCATTAAGACCTTGGGAGCTACAGATGCTATGCTCCACA 1314
Db 453 CCTGACGGGAGCCAGCTACTTTCATTAAGACCTTGGGAGCTACAGATGCTATGCTCCACA 394
QY 1315 GGAGGCCCCACGAATCCTGTCGGGCTGGAGGCTGTGAGAGGATGCTGGGGATAAGCCC 1374
Db 393 GGAGGCCCCACGAATCCTGTCGGGCTGGAGGCTGTGAGAGGATGCTGGGGATAAGCCC 334
QY 1375 TTAGGCACAGCTTAGACACCTTCCAAGAACCGGCCCGCTGATGCAAGATGGCAGATCT 1434
Db 333 TTAGGCACAGCTTAGACACCTTCCAAGAACCGGCCCGCTGATGCAAGATGGCAGATCT 274
QY 1435 GATACCCATTAGAGCCCCGAGAAATTCCTTCTGATCCAGTTTGCAGCAAAACCCACACA 1494
Db 273 GATACCCATTAGAGCCCCGAGAAATTCCTTCTGATCCAGTTTGCAGCAAAACCCACACA 214
QY 1495 CCCAGCTCACAGCAAAAAACAATGGACAGGCCCGCAGAGGCTGAAGCAAAACAGTGTCCCT 1554
Db 213 CCCAGCTCACAGCAAAAAACAATGGACAGGCCCGCAGAGGCTGAAGCAAAACAGTGTCCCT 154
QY 1555 TCTGGCTGTGTTGGAGCCTCCCGAGTAAACCACTATTTATTTTACCTCTTTCCAAACCT 1614
Db 153 TCTGGCTGTGTTGGAGCCTCCCGAGTAAACCACTATTTATTTTACCTCTTTCCAAACCT 94
QY 1615 GGAGCATTTATGCTAGGCTTGTCAAGAACTGTTCAAGTCCCTCTCTCTCAATAAAG 1674
Db 93 GGAGCATTTATGCTAGGCTTGTCAAGAACTGTTCAAGTCCCTCTCTCTCAATAAAG 34
QY 1675 CATCTTCAAGCTTGTAAAAAATAAANGA 1706
Db 33 CATCTTCAAGCTTGTAAAAAATAAANGA 2
```

RESULT 4

```
US-09-998-598-188
; Sequence 188, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 188
; LENGTH: 499
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-188
```

Query Match 28.5%; Score 487; DB 10; Length 499;
Best Local Similarity 99.8%; Pred. No. 2.1e-125;
Matches 498; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1187 CCTGAAGAGGTACAGACTGGCTGCAGGCAGAGAACTTCTCCACTSCACGGTGAAGACA 1246
Db 1 CCTGAAGAGGTACAGACTGGCTGCAGGCAGAGAACTTCTCCACTSCACGGTGAAGACA 60
QY 1247 CTTGGGTCCCTGACGGGAGCCAGCTACTTGCATAGAACCTGGGGAGCTACAGATGCTA 1306
Db 61 CTTGGGTCCCTGACGGGAGCCAGCTACTTGCATAGAACCTGGGGAGCTACAGATGCTA 120
QY 1307 TCTCCACAGGAGGCCCCAGCAATCTGTCCCGGCTGG-AGGCTGTCAGAAAGATGCTGGG 1365
Db 121 TGTCCACAGGAGGCCCCAGCAATCTGTCCCGGCTGGAGGCTGTCAAGAGATGCTGGG 180
QY 1366 GATAAGCCCTTAGGACCCAGCTTAGACACCTCTCAAGAACCCAGCCCGCTGATGCAAGAT 1425
Db 181 GATAAGCCCTTAGGACCCAGCTTAGACACCTCTCAAGAACCCAGCCCGCTGATGCAAGAT 240
QY 1426 GGCAGATCTGATACCGATTAGAGCCCGCGAGAAATCTCTCTTGGATCCAGTTTCAGCA 1485
Db 241 GGCAGATCTGATACCGATTAGAGCCCGCGAGAAATCTCTCTTGGATCCAGTTTCAGCA 300
QY 1486 AACCCACACCCAGCTACACAGCAAAACAAATGAGAGGCCAGAGGCTGAAGCAAAAC 1545
Db 301 AACCCACACCCAGCTACACAGCAAAACAAATGAGAGGCCAGAGGCTGAAGCAAAAC 360
QY 1546 AGTGTCCCTTCTGGCTGTGTGGAGCCTCCCCAGTAACCACTATTATTTTACCTCTTT 1605
Db 361 AGTGTCCCTTCTGGCTGTGTGGAGCCTCCCCAGTAACCACTATTATTTTACCTCTTT 420
QY 1606 CCCAAACCTGGAGCATTTATGCTTAGCTAGGCTGTCAAGATCTGTCAGTCCCTCTCTTCT 1665
Db 421 CCCAAACCTGGAGCATTTATGCTTAGGCTGTCAAGATCTGTCAGTCCCTCTCTTCT 480
QY 1666 CAATAAAGCATCTTCAAG 1684
Db 481 CAATAAAGCATCTTCAAG 499

RESULT 5

US-09-998-598-2183/c
; Sequence 2183, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: STOLK, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2183
; LENGTH: 447
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-2183

Query Match 24.0%; Score 411.2; DB 10; Length 447;
Best Local Similarity 94.5%; Pred. No. 2.2e-104;
Matches 415; Conservative 17; Mismatches 6; Indels 1; Gaps 1;

QY 1273 ACTTCGCATAAGACCTGGGGAGCTACAGATGCTATGTCCACAGGAGGCCCCACGAATCCT 1332
Db 447 ACTTGCATTAAGACCTGGGGAGCTACARATGCTATGTCCACAGGAGGCCCCACGAATCCT 388
QY 1333 GTCCCGGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCTTAGGCACCAAGCTTAGAC 1392
Db 387 GTCCCGGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCTTAGGCACCAAGCTTAGAC 328
QY 1393 ACCTCCAAAGAACAGGCCCGCTGTATGCAAGATGGCAGATCTGTATACCCATTAGAGCCCC 1452

Db 327 ACCTCCAAAGAACAGGCCCGCTGTATGCAAGATGGCAGATGTGATACCCATTAGAGCCCC 268
QY 1453 GAGAAATTCCTCTCTGGATCCAGTTTGCAGCAAAACCCACACCCAGC-TCACACAGCA 1511
Db 267 GAGAAATTCCTCTTGGATCCAGTTTGCAGCAAAACCCACACCCAGCSTTCACACAGCA 208
QY 1512 AAAACAATGGACAGGCCCCAGAGGCTGAAGCAAAACAGTCTCCCTTCTGGCTGTGTGGAGC 1571
Db 207 AAAACAATGGACAGGCCCCARAGGTTGAAGCAAAACAGTCTCCCTTCTGGCTGTGTGGAGC 148
QY 1572 CTCCCCAGTAAACACCTATTATTTTACCTCTTTTCCCAAACTGGAGCATTTATGCCCTAG 1631
Db 147 CTCCCCAGTAAACACCTATTATTTTACCTCTTTTCCCAAACTGGAGCATTTATGCCCTAG 88
QY 1632 GCTTGTCAAGAAATCTGTTTCAGTCCCTCTCTCAATAAAGAGCATCTTCAAGCTTGTAA 1691
Db 87 GCTTGTCAARAATCTGTTTCAGTCCCTCTCTTCAATAAAGAGCATCTTCAAGCTTGTAA 28
QY 1692 AAAAATAAATAANGATAAA 1710
Db 27 AAAAATAAATAAATAA 9
RESULT 6
US-09-880-107-290/c
; Sequence 290, Application US/09880107
; Patent No. US20020142981A1
; GENERAL INFORMATION:
; APPLICANT: Horne, Darci T.
; APPLICANT: Vockley, Joseph G.
; APPLICANT: Scherf, Uwe
; APPLICANT: Gene Logic, Inc.
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
; FILE REFERENCE: 44921-5028-WO
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 290
; LENGTH: 401
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 AAL34985
US-09-880-107-290

Query Match 22.7%; Score 389; DB 10; Length 401;
Best Local Similarity 99.8%; Pred. No. 3e-98;
Matches 400; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1259 ACGGGGACCCAGCTACTTCGCATAAGACCTGGGGAGCTACAGATGCTATGTCCA-CAGGA 1317
Db 401 ACGGGGACCCAGCTACTTCGCATAAGACCTGGGGAGCTACAGATGCTATGTCCACAGGA 342
QY 1318 GGCCCCAGCAATCCTGTCCGGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCTTA 1377
Db 341 GGCCCCAGCAATCCTGTCCGGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCTTA 282
QY 1378 GGCACACCTTAGACACCTTCCAAAGAACCCAGCCCGCTGTATGCAAGATGGCAGATCTGAT 1437
Db 281 GGCACACCTTAGACACCTTCCAAAGAACCCAGCCCGCTGTATGCAAGATGGCAGATCTGAT 222
QY 1438 ACCCAATTAGAGCCCGCAGAAATTCCTTCTTGGATCCCAAGTTTGCAGCAAAACCCACACC 1497
Db 221 ACCCAATTAGAGCCCGCAGAAATTCCTTCTTGGATCCCAAGTTTGCAGCAAAACCCACACC 162
QY 1498 CAGCTCACACAGCAAAACAAATGGACAGGCCCGCAGAGGCTGAAGCAAAACAGTGTCCCTTCT 1557

Db 161 CAGCTCACACGAAAAAACAATGACAGGCCAGAGGCTGAAGCAAAACAGTGTCCCTTCT 102
Qy 1558 GCGTGTGTGGAGCTCCCGAGTAACACACCTATTATTTACCTCTTTCCCAACCTGGA 1617
Db 101 GCGTGTGTGGAGCTCCCGAGTAACACACCTATTATTTACCTCTTTCCCAACCTGGA 42
Qy 1618 GCATTTATGCTAGGCTGTGCAAGAAATCTGTTCAGTCCCTC 1658
Db 41 GCATTTATGCTAGGCTGTGCAAGAAATCTGTTCAGTCCCTC 1
RESULT 7
US-09-764-868-522
; Sequence 522, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 522
; LENGTH: 572
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (3)
; OTHER INFORMATION: n equals a.t,g, or c
; NAME/KEY: SITE
; LOCATION: (27)
; OTHER INFORMATION: n equals a.t,g, or c
; NAME/KEY: SITE
; LOCATION: (516)
; OTHER INFORMATION: n equals a.t,g, or c
; NAME/KEY: SITE
; LOCATION: (552)
; OTHER INFORMATION: n equals a.t,g, or c
; NAME/KEY: SITE
; LOCATION: (558)
; OTHER INFORMATION: n equals a.t,g, or c
US-09-764-868-522

Query Match 19.6%; Score 335.4; DB 9; Length 572;
Best Local Similarity 97.7%; Pred. No. 2.7e-83;
Matches 339; Conservative 1; Mismatches 7; Indels 0; Gaps 0;
Qy 819 CTGTTTCCCTTCGGGGGAAGTCATAGTTAGGAGACCTCACACTTTCCCTCAGAGA 878
Db 165 CTCTGTTTCTTAGCGGGGAAGTCATAGTTAGGAGACCTCACACTTTCCCTCAGAGA 224
Qy 879 AGACACACACCATGACCCCTCAGCTGGGACCCCACTCCAGCCCTCCAGCCCAAC 938
Db 225 AGACACACACCATGACCCCTCAGCTGGGACCCCACTCCAGCCCTCCAGCCCAAC 284
Qy 939 CTGCCCCAGCCGCTCAAAATGCAAGTCTGTACGAGTTTGAAGCTAGGAACCCACGGG 998
Db 285 CTGCCCCAGCCGCTCAAAATGCAAGTCTGTACGAGTTTGAAGCTAGGAACCCACGGG 344
Qy 999 AACTGACTGTGGTCCAGGGAGAGAGCTGGAGTTCTGGACACAGCAAGCGTGTGGC 1058
Db 345 AACTGACTGTGGTCCAGGGAGAGAGCTGGAGTTCTGGRCACAGCAAGCGTGTGGC 404
Qy 1059 TGGTGAAGAATGAGCGGGGAGCGGCTACATTTCGAAGCAACATCCTGGAGCCCTAC 1118
Db 405 TGGTGAAGAATGAGCGGGGAGCGGCTACATTTCGAAGCAACATCCTGGAGCCCTAC 464

Qy 1119 AGCCGGGACCCCTGGGACCCAGGCGGCGAGTACCCCTCTCGGGTTCCA 1165
Db 465 AGCCGGGACCCCTGGGACCCAGGCGGCGAGTACCCCTCTCGGGTACTA 511
RESULT 8
US-09-764-868-1345/c
; Sequence 1345, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PTZ32
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1345
; LENGTH: 2643
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-1345
Query Match 18.9%; Score 324; DB 9; Length 2643;
Best Local Similarity 100.0%; Pred. No. 8.3e-80;
Matches 324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1366 GATAAGCCCTTAGGCACCCAGCTTAGACACCTTCAAGAACCCGCGCTGATGCAAGAT 1425
Db 324 GATAAGCCCTTAGGCACCCAGCTTAGACACCTTCAAGAACCCGCGCTGATGCAAGAT 265
Qy 1426 GGCAGATCTCATACCCATTAGAGCCCGAGAAATTCCTTCTGGATCCAGTTGCA 1485
Db 264 GGCAGATCTCATACCCATTAGAGCCCGAGAAATTCCTTCTGGATCCAGTTGCA 205
Qy 1486 AACCCACACCCAGCTCACACAGCAAAACAATGGACGCCAGAGGCTGAAGCAAAAC 1545
Db 204 AACCCACACCCAGCTCACACAGCAAAACAATGGACGCCAGAGGCTGAAGCAAAAC 145
Qy 1546 AGTGTCCTTCTGGCTGTGTGGAGCCTCCCGAGTAAACACCTATTTTACCTCTTT 1605
Db 144 AGTGTCCTTCTGGCTGTGTGGAGCCTCCCGAGTAAACACCTATTTTACCTCTTT 85
Qy 1606 CCCAAACCTGGAGCATTTATGCTAGGCTTGTCAAGAAATCTGTTCACTCCTCTCTCT 1665
Db 84 CCCAAACCTGGAGCATTTATGCTAGGCTTGTCAAGAAATCTGTTCACTCCTCTCTCT 25
Qy 1666 CAATAAAGCATCTTCAAGCTGT 1689
Db 24 CAATAAAGCATCTTCAAGCTGT 1
RESULT 9
US-09-925-299-321
; Sequence 321, Application US/09925299
; Patent No. US20020055627A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
; FILE REFERENCE: PA102
; CURRENT APPLICATION NUMBER: US/09/925,299
; CURRENT FILING DATE: 2001-08-10
; PRIOR APPLICATION NUMBER: PCT/US00/05883
; PRIOR FILING DATE: 2000-03-08
; PRIOR APPLICATION NUMBER: 60/124,270
; PRIOR FILING DATE: 1999-03-12
; NUMBER OF SEQ ID NOS: 1556
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 321
; LENGTH: 216
; TYPE: DNA
; ORGANISM: Homo sapiens

FEATURE:
NAME/KEY: misc_feature
LOCATION: (2)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (4)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (10)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (11)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (25)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (108)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (155)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (183)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (190)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (192)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (202)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-321

Query Match 7.0%; Score 119.2; DB 10; Length 216;
Best Local Similarity 84.5%; Pred. No. 1.6e-23;
Matches 142; Conservative 0; Mismatches 24; Indels 2; Gaps 1;
QY 1291 GGAGTCACAGTGTATCTCCACAGAGGCCCGCCAGGATCTGTCCCGGCTGGAGGCTGT 1350
Db 51 CGAGTACCGTCCGGAATTCGCGGTGCGACCCAGCGCTCC--GCCCGCTGGAGGCTGN 108
QY 1351 CAGAAGGATGCTGGGGTAAGCCCTTAGCGACCCAGCTTAGACACCTCCCAAGAACCGGCC 1410
Db 109 CAGAAGGATGCTGGGGTAAGCCCTTAGCGACCCAGCTTAGACACCTCCCAAGAACCGGCC 168
QY 1411 CGCGTGATGCAAGATGGCAGATCTGATACCCATTAGACCCCGGAGAT 1458
Db 169 CGCGTGATGCAAGATGGCAGATCTGATACCCATTAGACCCCGGAGAT 216

RESULT 10
US-09-925-299-578
Sequence 578, Application US/09925299
Patent No. US20020055627A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
CURRENT FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 578
LENGTH: 595
TYPE: DNA
ORGANISM: Homo sapiens

FEATURE:
NAME/KEY: misc_feature
LOCATION: (4)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (5)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (14)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (138)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (212)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (376)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (414)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (419)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (483)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (564)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (565)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (570)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (572)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc_feature
LOCATION: (576)
OTHER INFORMATION: n equals a,t,g, or c
US-09-925-299-578

Query Match 4.0%; Score 68.8; DB 10; Length 595;
Best Local Similarity 93.3%; Pred. No. 2.6e-09;
Matches 70; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 356 GGAAGCTGGAGAAGGCCCGCCAGGCAAGACCCAGGAGAAATTTGGGAAAAAAC 415
Db 1 GGANCTGGATTTGGCCCGCAAGACCCAGGAGAAATTTGGGAAAAAAC 60
QY 416 AAGGACCCAGGAGGT 430
Db 61 AAGGACCCAGGAGGT 75

RESULT 11
US-09-998-598-2415/c
Sequence 2415, Application US/09998598
Patent No. US20020150922A1
GENERAL INFORMATION:
APPLICANT: Stolk, John A.
APPLICANT: Xu, Jiangchun
APPLICANT: Chenault, Ruth A.
APPLICANT: Meagher, Madelein Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.561
CURRENT APPLICATION NUMBER: US/09/998,598
CURRENT FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 2606

```
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2415
; LENGTH: 245
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-2415

Query Match          3.5%; Score 59; DB 10; Length 245;
Best Local Similarity 96.0%; Pred. No. 8.6e-07;
Matches 72; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 GCACAGCGACTGAACACAGCCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAA---AG 57
    |||||
Db 75 GGCAGAGCGACTGAAGACCAAGCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAAGCAG 16
    |||||
QY 58 ACCTCGACTTGGAGG 72
    |||||
Db 15 ACCTCGACTTGGAGG 1

RESULT 12
US-09-764-868-99
; Sequence 99, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 99
; LENGTH: 1639
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-99

Query Match          3.3%; Score 56.4; DB 9; Length 1639;
Best Local Similarity 54.3%; Pred. No. 1.2e-05;
Matches 114; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 950 GCCTCGAAAATCAAGTCTTGACGAGTTTGAAGCTAGGAACCCAGCGGAAGTCACTGTG 1009
    |||
Db 186 GCCAAGTACGTCAAGATCCTGTATGACTTCACAGCCCGAAATGCCAACGAGCTATCGGTG 245
    |||
QY 1010 GTCCAGGGAGAGAGCTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGTGAAGAAT 1069
    |||
Db 246 CTCAAGGATGAGGTCTCTAGAGGTGCTGGAGGCGCGGCGGAGTGGTGAAGCTGCGCAGC 305
    |||
QY 1070 GAGGCGGGAGCGGCGCTACATTCCAAAGCAACATCTCTGAGCCCTACAGCGGGGACC 1129
    |||
Db 306 CGCAGCGGCGAGCGGGGTACGTGCCCTGCAACATCTAGGCGAGCGCGGACCGGAGGAC 365
    |||
QY 1130 CTTGGGACCCAGGCGGAGTCAACCTCTCGG 1159
    |||
Db 366 GCCGCGCCCGCTTCGAGCAGCGCGGTCTCAG 395
    |||

RESULT 13
US-09-764-868-512
; Sequence 512, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 512
; LENGTH: 1661
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-764-868-512

Query Match          3.3%; Score 56.4; DB 9; Length 1661;
Best Local Similarity 54.3%; Pred. No. 1.2e-05;
Matches 114; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 950 GCCTCGAAAATCAAGTCTTGACGAGTTTGAAGCTAGGAACCCAGCGGAAGTCACTGTG 1009
    |||
Db 175 GCGAAGTAGCTCAAGATCCTGTATGACTTCACAGCCCGAAATGCCAACGAGCTATCGGTG 234
    |||
QY 1010 GTCCAGGGAGAGAGCTGGAGGTTCTGGACCACAGCAACGCGTGGTGGTGAAGAAT 1069
    |||
Db 235 CTCAAGGATGAGGTCTCTAGAGGTGCTGGAGGAGCGCGGCGAGTGGTGAAGTGGCCAGC 294
    |||
QY 1070 GAGGCGGGAGCGGAGCGGCTACATTCCAAGCAACATCTCTGAGCCCTACAGCGGGGACC 1129
    |||
Db 295 CGCAGCGGCGCAGCGGGGTACGTGCCCTGCAACATCTCTAGGCGAGGCGCGGAGGAC 354
    |||
QY 1130 CTTGGGACCCAGGCGGAGTCAACCTCTCGG 1159
    |||
Db 355 GCCGCGCCCGCTTCGAGCAGCGCGGTCTCAG 384
    |||

RESULT 14
US-09-833-790-43
; Sequence 43, Application US/09833790
; Patent No. US20020068288A1
; GENERAL INFORMATION:
; APPLICANT: Lodes, Michael J.
; APPLICANT: Wang, Tongtong
; APPLICANT: Secrist, Heather
; APPLICANT: Mohamath, Raodoh
; APPLICANT: Indirias, Carol Y.
; APPLICANT: Fan, Liquon
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
; FILE REFERENCE: 210121.512
; CURRENT APPLICATION NUMBER: US/09/833,790
; CURRENT FILING DATE: 2001-04-11
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 514
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-833-790-43

Query Match          3.2%; Score 55; DB 10; Length 514;
Best Local Similarity 49.2%; Pred. No. 1.6e-05;
Matches 145; Conservative 0; Mismatches 150; Indels 0; Gaps 0;

QY 450 TTGACTGTCTCCAGAAGATCAAGTACAGCTTCAACCTCTCTGGGAAGGCTGGCCACCTGCG 509
    |||
Db 44 TTGACTGTCTCCAAAAGTTTAAACACGGATTAAACCTCTCTGGCCAACTGAAGTCTCATA 103
    |||
QY 510 TGAAGGAGACAAGTGCCCTGAGCTGCTGTACATCCTCTTCAAGTCCCTTGAACCTTCATCC 569
    |||
Db 104 TTCAGAAATCCTAGTGTGTCAGATTGGTTTCACTTTTGTACTCCATTAATAATATGGTGG 163
    |||
QY 570 TGGCCAGGTGCCCTGAGGCTGGCCTAGCAGCCCAAGTGTATCAACCTCTCTCACCCCTCCTCACCCCTA 629
    |||
Db 164 TCGAGCAACAGAGGAGTCTGTAACCTAGCCAGTTCAGTACTTAGTCCCTCTATTGAATAAG 223
    |||
QY 630 AAGCTATCAACCTGTCTACAGTCTCTGTCTAAGCCCACTCAGAGTAACTTTGGATGGGT 689
    |||
Db 224 ACACAATTGATTCTTAAATTTATCTGCAATGGTGTGAACGGCAGCTGTGGATGTCAAT 283
    |||
QY 690 TGGGCGCCAGCTGGACCACTAGCCGGCGGCGGAGTGGACAGCGCATGAGCCCTGCC 744
    |||
```

```

search completed: F
Job time : 100 secs

```

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 16:13:42 ; Search time 18.5 Seconds
(without alignments)
7667.516 Million cell updates/sec

Title: US-09-762-021a-1
Perfect score: 3127
Sequence: 1 ggcagagcagctgaagacca.....aaaaaaaaaataangataaa 1710

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09762021/runat_24022003_153038_7545/app_query.fasta_1.1863
-DB=SwissProt_40 -QFMT=fastan -SUFFIX=rsf -MINMATCH=0.1 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45
-DOALIGN=200 -THR_SCORE=ptc -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09762021@cgn_1.1.12 -runat_24022003_153038_7545 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	558.5	17.9	821	1 EPS8_MOUSE	Q08509 mus musculus
2	550	17.6	822	1 EPS8_HUMAN	Q12929 homo sapien
3	217	6.9	1454	1 CA13_MOUSE	P08121 mus musculus
4	200.5	6.4	1456	1 CA13_HUMAN	P02461 homo sapien
5	196.5	6.3	2161	1 SHK1_HUMAN	Q9y566 homo sapien
6	192	6.1	660	1 YHL1_EBV	P03181 epstein-bar
7	192	6.1	2167	1 SHK1_RAT	Q9wv48 rattus norv
8	186.5	6.0	1083	1 T2D3_HUMAN	O00268 homo sapien
9	185	5.9	963	1 YQ36_CAEEL	Q09457 caenorhabdi
10	184.5	5.9	1049	1 SHK1_BOVIN	P04258 bos taurus
11	184.5	5.9	1183	1 DRPL_RAT	P54258 rattus norv
12	184	5.9	2944	1 CA17_HUMAN	Q02388 homo sapien
13	183.5	5.9	2774	1 MAPA_RAT	P34926 rattus norv
14	181	5.8	1453	1 CA11_CHICK	P02457 gallus gall
15	180	5.8	1453	1 CA11_MOUSE	P11087 mus musculus
16	177.5	5.7	547	1 Z213_HUMAN	O14771 homo sapien
17	177	5.7	1729	1 TABP_HUMAN	Q9c0c2 homo sapien
18	174.5	5.6	1454	1 CA11_HUMAN	P02452 homo sapien

19	172.5	5.5	1185	1 DRPL_HUMAN	P54259 homo sapien
20	172	5.5	707	1 SFPO_HUMAN	P23246 homo sapien
21	171.5	5.5	1460	1 CA11_CANFA	Q9x517 canis fami
22	170	5.4	1516	1 CA1H_HUMAN	P39060 homo sapien
23	170	5.4	3530	1 MY15_HUMAN	Q9ukn7 homo sapien
24	168.5	5.4	487	1 EBN2_EBV	P12978 epstein-bar
25	168.5	5.4	671	1 VINE_HUMAN	O60504 homo sapien
26	168.5	5.4	2142	1 BAT2_HUMAN	P48634 homo sapien
27	166	5.3	1418	1 CA12_HUMAN	P02458 homo sapien
28	165	5.3	779	1 CA11_BOVIN	P02453 bos taurus
29	163.5	5.2	1336	1 W146_HUMAN	Q9c0j8 homo sapien
30	163.5	5.2	1459	1 CA12_MOUSE	P28481 mus musculu
31	163	5.2	680	1 CA1A_MOUSE	Q05306 mus musculu
32	162.5	5.2	636	1 CA13_RAT	P13941 rattus norv
33	162.5	5.2	1362	1 CA21_CHICK	P02467 gallus gall
34	162.5	5.2	3149	1 TEGU_EBV	P03186 epstein-bar
35	161.5	5.2	1804	1 CA1B_MOUSE	Q61245 mus musculu
36	161.5	5.2	3511	1 MY15_MOUSE	Q9qz24 mus musculu
37	161	5.1	699	1 VGLG_HSV2H	P13290 herpes simp
38	160.5	5.1	1364	1 CA21_BOVIN	P02465 bos taurus
39	160	5.1	1262	1 CA13_CHICK	P12105 gallus gall
40	158	5.1	867	1 SSPQ_BOVIN	P98167 bos taurus
41	158	5.1	1372	1 CA21_RAT	P02466 rattus norv
42	158	5.1	1838	1 CA15_HUMAN	P20908 homo sapien
43	157.5	5.0	1366	1 CA21_CANFA	O46392 canis fami
44	157	5.0	467	1 CBPA_DICDI	P35085 dictyosteli
45	157	5.0	1736	1 CA2B_HUMAN	P13942 homo sapien

ALIGNMENTS

RESULT 1	
EPS8_MOUSE	
ID	EPS8_MOUSE
AC	Q08509; STANDARD; PRT; 821 AA.
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	16-OCT-2001 (Rel. 40, Last annotation update)
DE	Epidermal growth factor receptor kinase substrate EPS8.
CN	EPS8.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX	NCBI_TaxID=10090;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=94008987; PubMed=8404850;
RA	Fazioli F., Minichiello L., Matoska V., Castagnino P., Mixi T.,
RA	Wong W.T., di Fiore P.P.;
RT	"Eps8, a substrate for the epidermal growth factor receptor kinase,
RT	enhances EGF-dependent mitogenic signals.";
RL	EMBO J. 12:3799-3808(1993).
RN	[2]
RP	X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS) OF 532-591.
RX	MEDLINE=97448677; PubMed=9303002;
RA	Kishan K.V.R., Scita G., Wong W.T., di Fiore P.P., Newcomer M.E.;
RT	"The SH3 domain of Eps8 exists as a novel intertwined dimer.";
RL	Nat. Struct. Biol. 4:739-743(1997).
CC	-!- FUNCTION: UPON BINDING TO EGF RECEPTOR ENHANCES EGF-DEPENDENT
CC	MITOGENIC SIGNALS. CAN BIND MULTIPLE CELLULAR TARGETS.
CC	-!- PTM: PHOSPHORYLATED BY SEVERAL RECEPTOR TYROSINE KINASES.
CC	-!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC	-!- SIMILARITY: CONTAINS 1 PH DOMAIN. IT IS SPLIT IN TWO PARTS.
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; L21671; AAA16358.1; -

DR PDB: 1AOJ; 08-JUL-98.
DR MGD; MGI:104684; Eps8.
DR InterPro: IPR000050; PID_domain.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00018; SH3; 1.
DR ProDom: PD000066; SH3; 1.
DR SMART: SM00462; PTB; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PS00002; SH3; 1.
KW SH3 domain; Phosphorylation; 3D-structure.
FT DOMAIN 69 129
FT DOMAIN 210 213
FT DOMAIN 322 325
FT DOMAIN 381 414
FT DOMAIN 421 440
FT DOMAIN 532 591
FT DOMAIN 620 650
FT DOMAIN 658 663
FT DOMAIN 821 AA; 91738 MW; 698959DD22D910C CRC64;
SQ SEQUENCE 821 AA; 91738 MW; 698959DD22D910C CRC64;
Alignment Scores:
Pred. No.: 5, 55e-25 Length: 821
Score: 558.50 Matches: 158
Percent Similarity: 38.82% Conservative: 78
Best local Similarity: 25.99% Mismatches: 171
Query Match: 17.86% Indels: 201
DB: 1 Gaps: 17
US-09-762-021a-1 (1-1710) x EPS8_MOUSE (1-821)
QY 101 GGGCCTGTATGAAAGCGCGCTCCCTATGAGCAGCGCAGCTATCTGGAGCGGGGATC 160
DB 190 GlyLysGlnLysArgProGluAlaLeuArgMetIleAlaLysAlaAspProGlyIle 209
QY 161 CCTCCAGACAGCCACCAGGAGCCCTAGACAGCCTCCACCACATCCCAAGGCC 220
DB 210 ProProPro-----ProArgAlaProAlaPro 218
QY 221 CTGCCAGCGCACACC-----AGTGCC 241
DB 219 ValProProGlyThrValThrGlnValAspValArgSerArgValAlaAlaTrpSerAla 238
QY 242 CGAGAACCAAGTCTTACTCTGCTCTCCAGCGGCTCTCTTCCCGAGGAC--- 298
DB 239 TrpAlaAlaAspGlnGlyAspPheGluLysProArgGlnTyrHisGluGlnGluThr 258
QY 299 CCAGAG-----AGGACGAGGAGTCTGCAACCATGCTCTAAGG 337
DB 259 ProGluMetMetAlaAlaArgIleAspArgValGlnIleLeuAsnHisIleLeuAsp 278
QY 338 GACATTGAGCTTCTATGGAAAGCTGGAGAGGCCAG-----GCA 379
DB 279 AspIleGluPhePheIleThrLysLeuGlnLysAlaAlaGluAlaPheSerGluLeuSer 298
QY 380 AGACACAGCAGGAGAAATTTGGGAAAAAACAAGGACCGAGGAGGTCTACC--- 436
DB 299 LysArgLysLysSerLysArgLysGlyProGlyGluGlyValLeuThrLeu 318
QY 437 -----CAGGCACAGTACATTGACTGCTTCCCAAGATCAAGTACAGC 478
DB 319 ArgAlaLysProProProProAspGluPheValAspCysPheGlnLysPheLysHisGly 338
QY 479 TTTCAACCTCTGGGAGGCTGGCCACCTGGCTGAAGGACAGCAAGTGCCTGAGCTCTCTA 538
DB 339 PheAsnLeuLeuAlaLysLysSerHisIleGlnAsnProSerAlaSerAspLeuVal 358
QY 539 CACATCTCTTCAAGTCCCTGAACCTTCATCTCGGCCAGGTCGCCCTGAGGCTGGCTAGCA 598
DB 359 HisPheLeuPheThrProLeuAsnMetValValGlnAlaThrGlyGlyProGluLeuAla 378
QY 599 GCCCAAGTATCTACCCCTCTCTACCCCTAAAGCTATCAACCTGCTACAGTCTGTCTA 658
DB 379 SerSerValLeuSerProLeuLeuThrLysAspThrValAspPheLeuAsnTyrThrAla 398

QY 659 AGCCCACTGAGAGTAACCTTTGGATGGGTTGGCCAGCCTGGACCACCTAGCGGGCC 718
DB 399 ThrAlaGluAlaArgLysLeuTrpMetSerLeuGlyAspSerTrpValLysValArgAla 418
QY 719 GACTGGACAGCGATGAGCCCTG---CCCTACCAACCCACATTTCTCAGATGACTGGCAA 775
DB 419 GluTrpProLysGluGlnPheIleProProTrpValProArgPheArgAsnGlyTrpGlu 438
QY 776 CTTCCA-----GAGCCCTCCAGCCCAAGCACCCTTAGGTAGTACCAGAC 817
DB 439 ProProMetLeuAsnPheMetGlyAlaProThrGluGlnAspMet-----TyrGlnLeu 456
QY 818 CCGTGTTCCTCGCGGGGAAGTCAT---AGTTTAGGAGCAGCCTCACACTTCTCCTCAG 874
DB 457 AlaGluSerValAlaAsnAlaGluHisGlnArgLysGlnAspSerLysArgLeuSerThr 476
QY 875 GAGAAGACACAAACCATGACCTCAGCCTGGGAC----- 910
DB 477 GluHisSerAsnValSerAspTyrProProAlaAspGlyTyrAlaTyrSerSerMet 496
QY 911 -----CCC 913
DB 497 TyrHisArgGlyProHisAlaAspHisGlyGluAlaAlaMetProPheLysSerThrPro 516
QY 914 AAC-----TCCAGGCGCTCCAGGCCCAACCTGCCAGCAGCCCTGAAATGCAA 964
DB 517 AsnHisGlnValAspArgAsnTyrAspAlaValLysThrGlnProLysLysTyrAlaLys 536
QY 965 GTCTGTGTACAGTTGAAGCTAGGAACCCCGGAACTAGCTGTGTCCAGGGAGAGAG 1024
DB 537 SerLysTyrAspPheValAlaArgAsnSerSerGluLeuSerValMetLysAspVal 556
QY 1025 CTGGAGTTCTCGACCAACAGCAGCGGTGGTGGTGGTGAAGAATCAGCGGGGAGG 1084
DB 557 LeuGluIleLeuAspAspArgArgGlnTrpLysValArgAlaSerGlyAspSer 576
QY 1085 GGCTACATTCCAAGCACATCTGGAG----- 1111
DB 577 GlyPheValProAsnAsnIleLeuAspIleMetArgThrProGluSerGlyValGlyArg 596
QY 1111 ----- 1111
DB 597 AlaAspProProTyrThrHisThrIleGlnLysGlnArgThrGluTyrGlyLeuArgSer 616
QY 1112 -----CCCCCTA 1117
DB 617 AlaAspThrProSerAlaProSerProProThrProAlaProValProValProLeu 636
QY 1118 CAGCGGGGAGCCCTGGG----- 1135
DB 637 ProProSerValProAlaProValSerValProLysValProAlaAspValThrArgGln 656
QY 1135 ----- 1135
DB 657 AsnSerSerSerSerAspSerGlyGlySerIleValArgAspSerGlnArgTyrLysGln 676
QY 1135 ----- 1135
DB 677 LeuProValAspArgArgLysSerGlnMetGluGluValGlnAspGluLeuPheGlnArg 696
QY 1136 ---ACCCAGGCGCAGTCCACCTCTCGG-----GTTCCA 1165
DB 697 LeuThrIleGlyArgSerAlaAlaGlnArgLysPheHisValProArgGlnAsnValPro 716
QY 1166 ATGCTTCGACTTAGC-----TCGAGGCGCTGAAGGTCACAGACTGCTGCAGGCAG 1219
DB 717 ValIleAsnIleThrTyrAspSerProGluGluValLysThrTrpLeuGlnSerLys 736
QY 1220 AACTTCTCCACTGCCCGTGGAGCAGTCTGGTCCCTCGCGGGGAGCGCAGCTACTCGC 1279
DB 737 GlyPheAsnProValThrValAsnSerLeuGlyValLeuAsnGlyAlaGlnLeuPheSer 756


```
QY 932 -----CCCAAACTGCCCGCCAGCCAGCCCTGAAATGCAAGTC 967
DB 519 ArgHisIleAspArgAsnTyrGluProLeuLysThrGlnProLysLysTyrAlaLysSer 538
QY 968 TTGTACGAGTTTGAAGCTAGGAAACCCAGGAGTCTGCTGCTCCAGGGAGAGAGCTG 1027
DB 539 LysTyrAspPheValAlaArgAsnSerGluLeuSerValLeuLysAspIleLeu 558
QY 1028 GAGGTCTCGACACAGCAAGCGGTGGTGGTGAAGATGAGCGGAGCGAGCGCGC 1087
DB 559 GluIleLeuAspArgLysGlnThrPrpLysValArgAsnAlaSerGlyAspSerGly 578
QY 1088 TACATTCGAAGCAACACTCTGGAG----- 1111
DB 579 PheValProAsnAsnIleLeuAspIleValArgProProGluSerGlyLeuGlyArgAla 598
QY 1111 ----- 1111
DB 599 AspProProTyrThrHisThrIleGlnLysGlnArgMetGluTyrGlyProArgProAla 618
QY 1112 -----CCCTACAG 1120
DB 619 AspThrProAlaProSerProProThrProAlaProValProValProLeuPro 638
QY 1121 CCGGGAGCCCTCTGG----- 1135
DB 639 ProSerThrProAlaProValProValSerLysValProAlaAsnIleThrArgGlnAsn 658
QY 1135 ----- 1135
DB 659 SerSerSerAspSerGlyGlySerIleValArgAspSerGlnArgHisLysGlnLeu 678
QY 1135 ----- 1135
DB 679 ProValAspArgArgLysSerGlnMetGluGluValGlnAspGluLeuIleHisArgLeu 698
QY 1136 ACCCAGGGCCAGTCACTCCCTCTCGG-----GTTCCAATG 1168
DB 699 ThrIleGlyArgSerAlaAlaGlnLysLysPheHisValProArgGlnAsnValProVal 718
QY 1169 CTTCACTTAGC-----TCGAGGCTGAAGAGTCACAGACTGCTGCAGCAGAGAAC 1222
DB 719 IleAsnIleThrTyrAspSerThrProGluAspValLysThrTrpLeuGlnSerLysGly 738
QY 1223 TTCTCCACTGCCAGCTGAGGACACTGGGTCCCTGACGGGAGCCAGCTACTTCGCATA 1282
DB 739 PheAsnProValThrValAsnSerLeuGlyValLeuAsnGlyAlaGlnLeuPheSerLeu 758
QY 1283 AGACCTGGGAGCTACAGTGTATGTCACAGAGGCGCCACGAATCCTGTCCCGGCTG 1342
DB 759 AsnLysAspGluLeuArgThrValCysPro---GluGlyAlaArgValTyrSerGlnIle 777
QY 1343 GAGGCTGTGAGAGGATGCTG 1363
DB 778 ThrValGlnLysAlaAlaLeu 784

RESULT 3
CAL3_MOUSE
ID CAL3_MOUSE STANDARD; PRT: 1464 AA.
AC P08121; Q61429; Q9CRN7;
DT 01-AUG-1988 (Rel. 08, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RC STRAIN=C57BL/6 X DBA; TISSUE=Embryo;
```

```
RX MEDLINE=95011609; PubMed=7926795;
RA "Toman D., de Crombrughe B.;
RT "The mouse type-III procollagen-encoding gene: genomic cloning and
RL complete DNA sequence.";
RN Gene 147:161-168(1994).
[2]
RP SEQUENCE OF 1-488 FROM N.A.
RX MEDLINE=88167858; PubMed=3443309;
RA Wood L., Theriault N., Vogeli G.;
RT "Complete nucleotide sequence of the N-terminal domains of the murine
RL alpha-1 type-III collagen chain.";
RN Gene 61:225-230(1987).
[3]
RP SEQUENCE OF 1-28 FROM N.A.
RX MEDLINE=85131189; PubMed=3972847;
RA Liau G., Mudryj M., de Crombrughe B.;
RT "Identification of the promoter and first exon of the mouse alpha 1
RL (III) collagen gene.";
RN J. Biol. Chem. 260:3773-3777(1985).
[4]
RP SEQUENCE OF 810-1464 FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Embryonic head;
MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Stauber P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Suzuki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RN Nature 409:685-690(2001).
[5]
RP SEQUENCE OF 1442-1464 FROM N.A.
RX STRAIN=C57BL/6;
MEDLINE=91274355; PubMed=2054384;
RA Metsaeranta M., Toman D., de Crombrughe B., Vuorio E.;
RT "Specific hybridization probes for mouse type I, II, III and IX
RL collagen mRNAs.";
RN Biochim. Biophys. Acta 1089:241-243(1991).
CC -1- FUNCTION: COLLAGEN TYPE III OCCURS IN MOST SOFT CONNECTIVE TISSUES
CC ALONG WITH TYPE I COLLAGEN.
CC -1- SUBUNIT: TRIMERS OF IDENTICAL ALPHA 1(III) CHAINS. THE CHAINS ARE
CC LINKED TO EACH OTHER BY INTERCHAIN DISULFIDE BONDS. TRIMERS ARE
CC ALSO CROSS-LINKED VIA HYDROXYLYSINES.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
CC UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC O-LINKED GLYCAN CONSISTS OF GLC-GAL DISACCHARIDE (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X52046; CAA36279.1; -
DR EMBL; M18933; AAA37338.1; -
DR EMBL; K03037; -; NOT_ANNOTATED_CDS.
DR EMBL; AK019448; BAB31724.1; -
```


DR EMBL; X57983; CAA41048.1; -.
 DR PIR; A22287; A22287.
 DR PIR; A27353; A27353.
 DR PIR; S16373; S16373.
 DR MGD; MGI:88453; Col3a1.
 DR InterPro; IPR000087; Collagen.
 DR InterPro; IPR000885; Fib_collagen_C.
 DR InterPro; IPR001007; VWF_C.
 DR Pfam; PF01391; Collagen; 18.
 DR Pfam; PF01410; COLFI; 1.
 DR ProDom; PD000007; Collagen; 1.
 DR ProDom; PD002078; Fib_collagen_C; 1.
 DR SMART; SM00038; COLFI; 1.
 DR SMART; SM00214; VWC; 1.
 DR PROSITE; PS01208; VWF; 1.
 KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
 KW Glycoprotein; Collagen; Signal.
 FT SIGNAL 1 23 BY SIMILARITY.
 FT PROPEP 24 154 AMINO-TERMINAL PROPEPTIDE.
 FT CHAIN 155 1203 COLLAGEN ALPHA 1(III) CHAIN.
 FT PROPEP 1204 1464 CARBOXYL-TERMINAL PROPEPTIDE.
 FT DOMAIN 31 90
 FT DOMAIN 155 169 NONHELICAL REGION (N-TERMINAL).
 FT DOMAIN 170 1195 TRIPLE-HELICAL REGION.
 FT DOMAIN 1196 1464 COLLAGEN ALPHA 1(III) CHAIN.
 FT CARBOHYD 262 262 O-LINKED (GAL. . .) (BY SIMILARITY).
 FT MOD_RES 262 262 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 283 283 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 859 859 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 976 976 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 1093 1093 HYDROXYLATION (BY SIMILARITY).
 FT MOD_RES 1105 1105 HYDROXYLATION (BY SIMILARITY).
 FT DISULFD 1195 1195 INTERCHAIN (BY SIMILARITY).
 FT DISULFD 1196 1196 INTERCHAIN (BY SIMILARITY).
 SQ SEQUENCE 1464 AA; 138944 MW; 2104EC27A886090B CRC64;

Alignment Scores:
 Pred. No.: 1,75e-05 Length: 1464
 Score: 217.00 Matches: 156
 Percent Similarity: 31.04% Conservative: 38
 Best Local Similarity: 24.96% Mismatches: 218
 Query Match: 6.94% Indels: 213
 DB: 1 Gaps: 34

US-09-762-021A-1 (1-1710) x CA13_MOUSE (1-1464)

Qy 10 ACTGAAGACCCGTCGACAA---GGCTCTGAGGAAGAGCTGGAGCAAAAGACCTCGACT 66
 Db 645 ThrGlyGlyProProGlyGluAsnGlyLysProGlyGluProGlyProLysGlyGluVal 664
 Qy 67 TGGAGGCTTCAGCCAGGCCAGGACAGATGGAGGGGGCTGCTATGGAAGGCCCTCCC 126
 Db 665 GlyAlaProGlyAlaProGlyGlyLysGlyAspSerGlyAlaProGlyGluArgGlyPro 684
 Qy 127 TATGGAGCA-----GGCAGCGTATCTGGAGCGGGGATCCCTCCAGAACAGCC 174
 Db 685 ProGlyThrAlaGlyLeuProGlyAlaArgGlyGlyAlaGlyProProGly----- 701
 Qy 175 CCACAGAGGACCCWAGA-----GCACAGCGTCCACCATCCCAAGGCCCTGCCACG 228
 Db 702 ---ProGlyGlyGlyLysGlyProAlaGlyProProGlyProProGlyAlaSerGlySer 720
 Qy 229 CCACAGCAGTCCCGAGCAACCAAGTGCTTTACTCTGCTCTCCAGCGCTCTCTTC 288
 Db 721 Pro-----GlyLeuGlnGly 725
 Qy 289 CCCCGAGGACCCAGAGGGAGGAAAGTGTGAACCATGCTTAAGGGACATTGAGCT 348
 Db 726 MetProGlyGluArgGlyGlyProGlySer-----ProGlyProLysGlyGlyLysGly 743
 Qy 349 GTTCATGGGAAGCTGGA---GAAGGC---CCAGCAAGACACGACGAGGAAGAAATT 402
 Db 744 GluProGlyGlyAlaGlyAlaAspGlyValProGlyLysAsp----- 757

Qy 403 TGGGAAAAAACAAGGACCA---GGGAGGTCTCACCAGGCACAGTACATTGACTGCT 459
 Db 758 -----GlyProArgGlyProAlaGlyPro----- 765
 Qy 460 CCAGAAGATCAAGTACAGCTTCAACCTCCTGGGAGGCTGGCCACCTGGCTGAAGGAGAC 519
 Db 766 -----IleGlyProProGlyProAlaGlyGln-----ProGlyAsp 777
 Qy 520 AAGTGCCCTCAGCTGCTACACATCTTCAAGTCCCTGAACCTTCATCTCCTGGCAGGTG 579
 Db 778 LysGlyGluGlyGlySer-----ProGlyLeu---ProGlyLeuAla 790
 Qy 580 ---CCCTGAGGCTGGCTAGCAGCCCAAGTAGTATCATCCCTCCCTCACCCCTAAAGCTAT 636
 Db 791 GlyProArgGlyGlyProGlyGluArgGlyGluHisGlyProProGlyProAlaGlyPhe 810
 Qy 637 CAACCTGCTACAGTCTGTCTAAGCCCACTGAGAGTAACCTTTGGATGGGTTG----- 691
 Db 811 ProGlyAlaProGlyGlnAsn-GlyGluProGlyAlaLysGlyGluArgGlyAlaProGly 830
 Qy 692 -----GGCCAGCTGGACCACTAGCCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 735
 Db 830 yGluLysGlyGluGlyGlyPro-ProGlyPro---AlaGlyProThrGlySerSerGlyP 849
 Qy 736 GCCCTGCTCCACCAACCCACATCTCAGATGACTGGCAACTTCCAGAGCCCTCCAGCCA 795
 Db 849 roAlaGlyProProGlyProGlnGly---ValLysGlyGluArgGlySerPro----- 865
 Qy 796 AGCACCTTAGGATACAGGACCCCTGTTTCCCTTCGGCGGGGAGTATAGTTAGGGAG 855
 Db 866 -----GlyG 867
 Qy 856 CACCTCACACTTCTCTCAGGAGAGACACACAACCATCACCTCAGCCTGGGGACCCCA- 914
 Db 867 lyProGlyThr-----AlaGlyPheProG 875
 Qy 915 -----ACTCCAGGCGCTCCAGCCCAACCTGCCAGCCAGCCGCTGAAATGCA 963
 Db 875 lyGlyArgGlyLeuProProGlyProProGlyAsnAsn-----GlyAsnPro----- 889
 Qy 964 AGTCTGTACGAGTTTGAAGCTAGGACCCACGGGAACCTGCTGTGTCCAGGAGAGAA 1023
 Db 890 -----Gly-ProProGlyProSerGlyAlaProGlyLysAsp 901
 Qy 1024 -----GCTGGAGGTCTCGA---CCACAGCAAGCGGTGGTGGCTGGTGA 1065
 Db 902 GlyProProGlyProAlaGlyAsnSerGlySerProGlyAsnProGlyLeuAlaGlyPro 921
 Qy 1066 GAATGAGCGGGAGCGGCTACATTCCAAAGCAACATCTGGAGCGCTTACAGCCCGG 1125
 Db 922 Lys---GlyAspAlaGlyGlnProGlyGlyLysGlyProProGlyAla-----GlnGly 938
 Qy 1126 GACCCCTGGGACCCAGGCCAGTCAACCTCTCGGTTTCCAATGCTTCGACTTAGCTC--- 1182
 Db 939 ProProGlySerProGlyProLeuGlyLeuAlaGlyLeuThrGlyAlaArgGlyLeuAla 958
 Qy 1183 -----GAGGCTGAAGAGGTACAGACTGGCT 1209
 Db 959 GlyProProGlyMetProGlyProArgGlySerProGlyProGlnGlyLeuLysGlyGlu 978
 Qy 1210 GCAGGCAGAGAATCTTCCACTGCCAGCGGTGAGGACACTTGGTCCCT----- 1257
 Db 979 SerGlyLysProGlyAlaSerGlyHisAsnGlyGluArgGlyProProGlyProGlnGly 998
 Qy 1258 -----GAC 1260
 Db 999 LeuProGlyGlnProGlyThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAsp 1018
 Qy 1261 GGGGAGCCAGCTACTTCCATTAAGACCTGGGAGGCTACAGATGCTATGTCACAGCA--- 1317
 Db 1019 GlyGlnProGlyArgAspGlySerProGlyGlyLysGlyAspArgGlyGluAsnGlySer 1038

```

QY 1318 -----GGCCCCAGAAATCTGTCCCGGCTGGAGGC 1347
Db 1039 ProGlyAlaProGlyAlaProGlyHisProGlyProGlyProValGlyProSerGly 1058
QY 1348 TGTCAGAAGATCTGGGGATAG-----CCCTTAGGCACACAGCTTAGACACC 1395
Db 1059 LysSerGlyAspArgGlyGluThrGlyProAlaGlyProSerGlyAlaProGlyProAla 1078
QY 1396 TCACAAGAACAGCCCGCTGATGCAAGATGCGAGATCTGATACCATTCATAGAGCCCGAG 1455
Db 1079 GlyAlaArgGlyAlaProGlyProGlnGlyProArgGlyAspLysGlyGluThrGlyGlu 1098
QY 1456 AATTCCTCTCTGGATCCAG-----TTTGCAGCAAAACCCACACCCCGAGCTC 1503
Db 1099 ArgGlySerAsnGlyIleLysGlyHisArgGlyPheProGlyAsnProGlyPro----- 1116
QY 1504 ACACAGCAAAACATGGACGCCAGCGCTGAAGCAAGCAAGTGTCTCTTCGCTGT 1563
Db 1117 -----ProGlySerProGlyAlaAlaGlyHisGlnGlyAla 1128
QY 1564 GTTGAGCCTCCCGAGTAACACCATTTATTTTACCTCTTTCCCAAACTGGAGCATTT 1623
Db 1129 IleGlySer-----ProGlySerProGlyAlaAlaGlyHisGlnGlyAla 1135
QY 1624 ATGCCCTAGG 1632
Db 1136 GlyProArg 1138
RESULT 4
CA13_HUMAN
ID CA13_HUMAN STANDARD; PRT; 1466 AA.
AC P02461; Q15112;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(III) chain precursor.
GN COL3A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
[1]
RC TISSUE=Skin fibroblast;
RX MEDLINE=89350838; PubMed=2764886;
RA Ala-Kokko L., Kontusaari S., Baldwin C.T., Kuivaniemi H.,
RA Prockop D.J.;
RT "Structure of cDNA clones coding for the entire prepro alpha 1 (III)
RT chain of human type III procollagen. Differences in protein structure
RT from type I procollagen and conservation of codon preferences.";
RL Biochem. J. 260:509-516(1989).
[2]
RN
RP SEQUENCE OF 149-1225 FROM N.A.
RX MEDLINE=89386015; PubMed=2780304;
RA Janeczko R.A., Ramirez F.;
RT "Nucleotide and amino acid sequences of the entire human alpha 1
RT (III) collagen.";
RL Nucleic Acids Res. 17:6742-6742(1989).
[3]
RN
RP SEQUENCE OF 168-398.
RX MEDLINE=77134724; PubMed=557335;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of cyanogen
RT bromide peptides from the amino-terminal segment of type III collagen
RT of human liver.";
RL Biochemistry 16:1158-1164(1977).
[4]
RN
RP REVISIONS.
RA Seyer J.M.;
RL Submitted (DEC-1977) to the PIR data bank.
[5]
RN
RP SEQUENCE OF 399-727.
RX MEDLINE=79000343; PubMed=687591;

```

```

RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of five
RT consecutive CNBR peptides from type III collagen of human liver.";
RL Biochemistry 17:3404-3411(1978).
[6]
RN
RP SEQUENCE OF 728-964.
RX MEDLINE=80198282; PubMed=6246925;
RA Seyer J.M., Mainardi C., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of alpha 1
RT (III)-CB5 from type III collagen of human liver.";
RL Biochemistry 19:1583-1589(1980).
[7]
RN
RP SEQUENCE OF 950-1466 FROM N.A.
RX MEDLINE=88189827; PubMed=3357782;
RA Mankoo B.S., Dalgleish R.;
RT "Human pro alpha 1(III) collagen: cDNA sequence for the 3' end.";
RL Nucleic Acids Res. 16:2337-2337(1988).
[8]
RN
RP REVISION TO 1184.
RX MEDLINE=89098346; PubMed=3211760;
RA Molyneux K., Dalgleish R.;
RT "Human type III collagen 'variant' is a cDNA cloning artefact.";
RL Nucleic Acids Res. 16:11833-11833(1988).
[9]
RN
RP SEQUENCE OF 1065-1466 FROM N.A.
RX MEDLINE=85087944; PubMed=6096827;
RA Loidl H.R., Brinker J.M., May M., Pihlajaniemi T., Morrow S.,
RA Rosenbloom J., Myers J.C.;
RT "Molecular cloning and carboxyl-propeptide analysis of human type III
RT procollagen.";
RL Nucleic Acids Res. 12:9383-9394(1984).
[10]
RN
RP SEQUENCE OF 965-1200.
RX MEDLINE=81208139; PubMed=7016180;
RA Seyer J.M., Kang A.H.;
RT "Covalent structure of collagen: amino acid sequence of alpha
RT 1(III)-CB9 from type III collagen of human liver.";
RL Biochemistry 20:2621-2627(1981).
[11]
RN
RP SEQUENCE OF 1176-1466 FROM N.A.
RX MEDLINE=85157600; PubMed=2579949;
RA Chu M.-L., Weil D., de Wet W.J., Bernard M.P., Sippola M., Ramirez F.;
RT "Isolation of cDNA and genomic clones encoding human pro-alpha 1
RT (III) collagen. Partial characterization of the 3' end region of the
RT gene.";
RL J. Biol. Chem. 260:4357-4363(1985).
[12]
RN
RP SEQUENCE OF 1161-1200 FROM N.A.
RX MEDLINE=86187804; PubMed=3754462;
RA Miskulin M., Dalgleish R., Klueve-Beckerman B., Rennard S.I.,
RA Tolstoshev P., Brantly M., Crystal R.G.;
RT "Human type III collagen gene expression is coordinately modulated
RT with the type I collagen genes during fibroblast growth.";
RL Biochemistry 25:1408-1413(1986).
[13]
RN
RP SEQUENCE OF 1-170 FROM N.A.
RX MEDLINE=88303360; PubMed=3405773;
RA Toman D., Ricca G., de Crombrughe B.;
RT "Nucleotide sequence of a cDNA coding for the amino-terminal region
RT of human prepro alpha 1(III) collagen.";
RL Nucleic Acids Res. 16:7201-7201(1988).
[14]
RN
RP SEQUENCE OF 1-176 FROM N.A.
RX MEDLINE=8938752; PubMed=2777083;
RA Benson-Chanda V., Su M.W., Weil D., Chu M.-L., Ramirez F.;
RT "Cloning and analysis of the 5' portion of the human type-III
RT procollagen gene (COL3A1).";
RL Gene 78:255-265(1989).
[15]
RN
RP REVIEW ON VARIANTS.
RX MEDLINE=97255959; PubMed=9101290;
RA Kuivaniemi H., Tromp G., Prockop D.J.;

```



```
QY 742 GCCTTACCACCCACATTCCTCAGATGACTGGCAACTTCAGAGCCCT----- 788
DB 852 lyProGlyProGlnGly---ValysGlyGluArgGlySerProGlyGlyProGlyA 871
QY 789 -----
DB 871 laAlaGlyPheProGlyAlaArGlyLeuProGlyPro-ArgGlySerAsnGlyAsnPro 890
QY 814 GGACCTGTTTCCTTCGGCGGGAGTGCATAGTTAGGAGCACCCTCACACTTCCTCA 873
DB 891 GlyProGlyProSerGly-----SerPro 899
QY 874 GGAGAAGCACACCAACCATGACCTCA-----GCC 903
DB 900 GlyLysAspGlyProProGlyProAlaGlyAsnThrGlyAlaProGlySerProGlyVal 919
QY 904 TGGGGACCCCACTCCAGCCCTCCAGCCC-----CAAACC 939
DB 920 SerGlyProLysGlyAspAlaGlyGlnProGlyGluLysGlySerProGlyAlaGlnGly 939
QY 940 TGCCACGACGACCTCAAAATGCAAGTCTGTACGAGTTTGAAGCTAGGACCCAGGGA 999
DB 940 ProProGlyAlaProGlyProLeuGlyIleAlaGlyIleThrGlyAlaArGlyLeuAla 959
QY 1000 ACTGACTGTGTCGAGGAGAGAGAGCTGGAGGTTCTGGACACGACGAGCGGTGGTGC 1059
DB 960 GlyProGlyMetProGlyProArGlySerProGlyProGlnGly-----ValLys 977
QY 1060 GTTGAGAGTAGGC-----GGGACGGAGCGGCTACATTCGAAGCAACATCCTGAGGC 1113
DB 978 GlyGluSer---GlyLysProGlyAlaAsnGlyLeuSerGlyGluArgGlyProProGly 996
QY 1114 CCF-----ACAGCGGGGAGCCCC-----TGGGACCA 1140
DB 997 ProGlnGlyLeuProGlyLeuAlaGlyThrAlaGlyGluProGlyArgAspGlyAsnPro 1016
QY 1141 GGCCCACTCACCCTCGGTTCCAAATGCTTCGACTTAGCTCGAGCCCTGAAGAGCTAC 1200
DB 1017 GlySerAspGlyLeuProGlyArgAspGlySerProGlyGlyLysGlyAspArgGlyGlu 1036
QY 1201 AGACTGGCTGACGAGAGAACTTCTCACTGCCCGGTGAGGACACTGGGTCCCTGAC 1260
DB 1037 -----Asn 1037
QY 1261 GGGAGCCAGCTACTTTCGATAGACCTGGGAGCTACAGATGCTATGTCACAGAGGC 1320
DB 1038 GlySerPro-----GlyAlaProGlyAlaProGlyHisProGly 1050
QY 1321 CCACGAATCTGTCCTCCGCTGGAGCTGTCAAGAGGATGCTGGGGATAAG----- 1371
DB 1051 ProProGlyProValGlyProAlaGlyLysSerGlyAspArgGlyGluSerGlyProAla 1070
QY 1372 ---CCCTTAGGCACAGCTTAGACCTCAAGACCCGCGCTGATGCAAGATGCC 1428
DB 1071 GlyProAlaGlyAlaProGlyProAlaGlySerArgGlyAlaProGlyProGlnGlyPro 1090
QY 1429 AGATCTGATACCATAGAGCCCGAGAAATCTCTCTCTGGATCCAG----- 1476
DB 1091 ArgGlyAspLysGlyGluThrGlyGluArgGlyAlaAlaGlyIleLysGlyHisArgGly 1110
QY 1477 TTTCGACCAACCCC-----ACACCCAGCTCACACAGCAACAAACAAATGGACA 1524
DB 1111 PheProGlyAsnProGlyAlaProGlySerPro----- 1121
QY 1525 GGGCCAGAGCTCAACCAACAGCTGTC-----CCTCTGCTGCTGT 1563
DB 1122 GlyProAlaGlyGlnGlnGlyAlaIleGlySerProGlyProAlaGlyProArgGlyPro 1141
QY 1564 GTTGGAGCCTCC 1575
DB 1142 ValGlyProSer 1145
```

```
RESULT 5
ID SHKL_HUMAN STANDARD: PRT: 2161 AA.
AC Q9V566: Q9NYW9;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE SH3 and multiple ankyrin repeat domains protein 1 (Shank1)
DE (Somatostatin receptor interacting protein) (SSTR interacting protein)
DE (SSRIP).
GN SHANK1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), AND INTERACTION WITH SSTR2.
RC TISSUE=Fetal brain, Hippocampus, and Thalamus;
RA MEDLINE=20020275; PubMed=10551867;
RX Zitzer H., Hoenck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;
RT "Somatostatin receptor interacting protein defines a novel family of
multidomain proteins present in human and rodent brain.";
RL J. Biol. Chem. 274:32997-33001(1999).
RN [2]
RP REVIEW.
RX PubMed=10806096;
RA Sheng M., Kim E.;
RL "The Shank family of scaffold proteins.";
RL J. Cell Sci. 113:1851-1856(2000).
CC -!- FUNCTION: Seems to be an adapter protein in the postsynaptic
density (PSD) of excitatory synapses that interconnects receptors
of the postsynaptic membrane including NMDA-type and metabotropic
glutamate receptors via complexes with GKAP/PSD-95 and Homer,
respectively, and the actin-based cytoskeleton. May play a role in
the structural and functional organization of the dendritic spine
and synaptic junction.
CC -!- SUBUNIT: May homodimerize via its SAM domain (By similarity).
Interacts with SSTR2 C-terminus via the PDZ domain. Interacts with
SPAN1, Homer-1 and DLGAP1/GKAP isoforms 1 and 2 (By similarity).
Is part of a complex with DLG4/PSD-95 and DLGAP1/GKAP (By
similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of
neuronal cells (By similarity).
CC -!- ALTERNATIVE PRODUCTS: 3 isoforms; 1/a (shown here), 2/b and 3; are
produced by alternative splicing.
CC -!- TISSUE SPECIFICITY: Expressed in brain particularly in the
amygdala, hippocampus, substantia nigra and thalamus. Isoform 2
seems to be expressed ubiquitously.
CC -!- SIMILARITY: BELONGS TO THE SHANK FAMILY.
CC -!- SIMILARITY: CONTAINS 6 ANK REPEATS.
CC -!- SIMILARITY: CONTAINS 1 PDZ/DHR DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AF163302; AAD45121.1; -
DR EMBL: AF226728; AAF35887.1; -
DR HSP: P06241; 1SHF.
DR MIM: 604999; -
DR InterPro: IPR002110; ANK.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001660; SAM.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00023; ank; 7.
DR Pfam: PF00595; PDZ; 1.
DR Pfam: PF00536; SAM; 1.
```

DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00248; ANK; 3.
DR SMART; SM00228; PDZ; 1.
DR SMART; SM00454; SAM; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50088; ANK_REPEAT; 3.
DR PROSITE; PS50297; ANK_REPEAT; 1.
DR PROSITE; PS50105; SAM_DOMAIN; 1.
DR PROSITE; PS50002; SH3; 1.
KW ANK repeat; SH3 domain; Repeat; Alternative splicing.
FT REPEAT 212 245 ANK 1.
FT REPEAT 246 278 ANK 2.
FT REPEAT 279 312 ANK 3.
FT REPEAT 313 345 ANK 4.
FT REPEAT 346 378 ANK 5.
FT REPEAT 379 395 ANK 6.
FT DOMAIN 554 613 SH3.
FT DOMAIN 663 757 PDZ.
FT DOMAIN 2098 2161 SAM.
FT DOMAIN 1002 1007 POLY-HIS.
FT DOMAIN 1014 1019 POLY-HIS.
FT DOMAIN 1189 1195 POLY-GLY.
FT DOMAIN 1709 1717 POLY-GLY.
FT DOMAIN 1844 1854 POLY-PRO.
FT DOMAIN 1896 1902 POLY-GLY.
FT DOMAIN 1970 1979 POLY-SER.
FT VARSPPLIC 1 613 MISSING (IN ISOFORM 2).
FT VARSPPLIC 614 654 RSQESKRSRDKAKLFRHYTVGSYDSFSDAPSLMDGIGPG
FT FT -> MOLMALEQRFSGPLGGGQPLCLMMSSPLPPPHFSC
FT FT LPA (IN ISOFORM 2).
FT VARSPPLIC 646 654 MISSING (IN ISOFORM 3).
SQ SEQUENCE 2161 AA; 225019 MW; 5FEFC969CBE98701 CRC64;

Alignment Scores:

Pred. No.: 0.00026 Length: 2161
Score: 196.50 Matches: 174
Percent Similarity: 30.70% Conservative: 51
Best Local Similarity: 23.74% Mismatches: 265
Query Match: 6.28% Indels: 243
DB: 1 Gaps: 35

US-09-762-021A-1 (1-1710) x SHK1_HUMAN (1-2161)

QY 25 GCAGAGGCTCTGGAGAGAGCTGGAGCAAGACCTCGACTGGAGGCTTCAGCCAGG 84
DB 1319 AlaTyrGlyGlyGlyGlySerSerAlaPheThrSerPheLeuProProArgProLeu 1338
QY 85 CCAGGACAGATGAGGGGGCTGTATGGAAGGCGCTCCCTATGAGGAGCGG----- 138
DB 1339 ValHisProLeuThrGlyLysAlaLeuAspProHisSerProLeuGlyLeuAlaLeuAla 1358
QY 139 -----ACGCTATCTGGAGCGGGGATCCCTCCAGAACAGCC 174
DB 1359 AlaArgGluArgAlaLeuLysGluSerSerGlyGlyGlyAlaProGlnProProPro 1378
QY 175 CCACCAGAGACCTAGACGACACGCTCCACACATCCCAAGGCCCTGCCAGCCACAC 234
DB 1379 ArgProProSerProArg-TyrGluAlaProPro-----ProThrProHisHisH 1395
QY 235 CAGTGCCCGAGAACAGTGCCTTTACTGCTCTCCCTCCCAAGGGG---TCCTCTCCCC 291
DB 1395 sSer-----ProHisAlaHisGluProValLeuArgLeuTrpGlyAlaSerPr 1412
QY 292 CGAGGACCCAGAGAGGACGAGGAAGTGCCTGAACATGCTCCTAAGGACATTTGAGCTGT 351
DB 1412 oProAspProAlaArgArgGluLeuGlyTyrArgAlaGlyLeu-GlySerGlnGluLys 1432
QY 352 CATGGGAAAGCTGGAGAGGCCCGGCAAGACACGAGGAGAAATTTGGGAAAAA 411
DB 1432 er-----LeuProAlaSerProProAlaAlaArgArgSerLeu-----L 1445

QY 412 AAACAAGGACGAGGAGGTCTCACCCAGGACAGTACATTGCTTCAGAGATCAA 471
DB 1445 euHisArgLeuProProThrAlaProGlyValGlyProLeuGlnLeuGlyThrG 1465
QY 472 GTACA-----GCTTCAACCTCTGGGAAGGCTGCCACCTGGGTGAA 513
DB 1465 luProProAlaProHisProGlyValSerLysProTrpArgSerAlaAlaProGluGluP 1485
QY 514 GGAGACAAGTGCCCTG-----AGTCGTGTACA 540
DB 1485 roGluArgLeuProLeuHisValArgPheLeuGluAsnCysGlnProArgAlaProVal 1505
QY 541 CATCCCTCTTCAAGTCCCTGAACCTTCATCTGGCAGGTGCCCTGAGCTGCCCTAGCAGC 600
DB 1505 hrSerGlyArgGlyProProSerGluAspGlyProGlyVal-----Pro-ProProSer 1522
QY 601 CCAAGTGATCTACCCCTCTCTACCCCTAAAGCTATCAA----- 639
DB 1523 ProArgArgSerValProProSerProThrSerProArgAlaSerGluGluAsnGlyLeu 1542
QY 640 CCTGCTACAGTCTGTCTAAGCCACCTGAGAGTAACCTTTGGATGGGTGGGCCCGCAGC 699
DB 1543 ProLeuLeuValLeuProPro-----AlaProSer 1553
QY 700 CTGACCACTAGCCGGCGCAGCTGGACGCGATGAGCCCTGCCCTTACCAACCCACATT 759
DB 1554 ValAsp-ValGluAspGlyGluPheLeuPheValGluProLeuProProLeuGluPh 1573
QY 760 CTCAGATGACTGCAACTTCCAGAG----- 784
DB 1573 eSerAsnSerPheGluLysProGluSerProLeuThrProGlyProProHisProLeuPr 1593
QY 785 -----CCCTCCAGCAACACACCTTAGATACACGACCCCTGTTCCCTCGCGCGGG 837
DB 1593 oAspThrProAlaProAlaThrProLeuProProValProProAlaValAlaAlaAl 1613
QY 838 AAGTCATAGTTAGGGAGCAC-----TCACACTTTCTCCAGGAG----- 877
DB 1613 aProProThrLeuAspSerThrAlaSerSerLeuThrSerTyrAspSerGluValAlaTh 1633
QY 878 -AAGACACACACCATGACCTCAGCTGGGGGACCCCACTCCA---GGCCCTCCAGCC 933
DB 1633 rLeuThrGlnGlyAlaSerAlaAlaProGlyAspProHis-ProProGlyProProAlaP 1653
QY 934 CAACCTGCCAGCC-----AGCCCTGAAATGCAAGCTCTGTGACA 975
DB 1653 ro-AlaAlaProAlaProAlaAlaProGlnProGlyProAsp----- 1666
QY 976 GTTTGAAGCTAGGAACCCAGGAACTGACTGT----- 1008
DB 1667 -----ProProGlyThrAspSerGlyIleGluGluValAspSerArgSer 1682
QY 1008 ----- 1008
DB 1683 SerSerAspHisProLeuGluThrIleSerSerAlaSerThrLeuSerSerLeuSerAla 1702
QY 1009 ---GGTCCAGGAGAGAGCTGGAGTTCTGGACCACAGACGGGTGGTGGGTGAA 1065
DB 1703 GluGlyGlySerAlaGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 1722
QY 1066 GAATGAGCGGGGAGC-----GAGCGGCTACATTCCAAAGCAACATCTGGAGCCCTTACA 1119
DB 1723 GluLeuLeuAspThrTyrValAlaTyrLeuAspGlyGlnAlaPheGlyGlySerSerThr 1742
QY 1120 GCCGGGACCCCTGGGACCCAGGCGCCAGTCCCTCTCGGGTTCCCAATGCTCGACTTAG 1179
DB 1743 ProGlyProProProProGlnLeuMetThrProSerLysLeu----- 1757
QY 1180 CTCGAGGCTGAGAGGTCACAGCTGGCTGCGAGGACAGAACTTCTCCACTGCCACGGT 1239
DB 1758 -----ArgGlyArgAlaLeuGlyAlaSerGlyGlyLeuArg-----ProGly 1771
QY 1240 GAGGACACTTGGTCCCTGACGGGGAGCCAGCTACTTCG-----CATAAGACCTGG 1290

```
Db 1772 ProSerGlyLeuArgAspValThrProThrSerProThrValSerValThrGly 1791
QY 1291 GGAGCTACAGATGCTATG-::: ||||| ||||| |||||
Db 1792 AlaGlyThrAspGlyLeuAlaLeuAlaCysSerGlyProProThrAlaGlyVal 1811
QY 1312 ACAGGAGCCACCAAGT-::: ||||| ||||| |||||
Db 1812 AlaGlyGlyProValAlaValGluProGluValProValProThrAlaSer 1831
QY 1348 TGT-----CAGAGGATGCTGGGAT-::: ||||| ||||| |||||
Db 1832 CysLeuProArgLysLeuLeuProTirpGluGluGlyProGlyProProProLeu 1851
QY 1369 -----AAGCCCTTAGCCACAGCTTAGAC-::: ||||| ||||| |||||
Db 1852 ProGlyProLeuAlaGlnProGlnAlaSerHisGlyValAlaSerIle 1871
QY 1396 TCACAAACAGCCCGCTGATGCAAGATGCAGATCT-::: ||||| ||||| |||||
Db 1872 SerGluLeuSerSerLysLeuGlnPheGlyGlySerSerAlaAlaGlyAlaLeu 1891
QY 1435 -----GATACCCATTAGAC-::: ||||| ||||| |||||
Db 1892 ProTirpAlaArgGlySerGlyGlyGlyAspSerHisGlyValAlaSerTyr 1911
QY 1450 ---CCGAGAAATTCCTCT-::: ||||| ||||| |||||
Db 1912 ValProGluArgThrSerSerLeuGlnArgGlnLeuSerAspSerGlnSer 1931
QY 1477 TTTGAGCAAAACCCACAGCTCACAGCAAAACATGAGCCAGAGGCT 1536
Db 1932 LeuLeuSerLysProValSerSerLeuPheGln-----AsnTirp-----ProLysPro 1947
QY 1537 GAAGCAACAGTGTCTTGGCTGTGTGGAGCTCCAGTAACACCTATTATT 1596
Db 1948 ProLeuProProLeuProThrGly---ThrGlyValSerProThrAlaAla 1633
QY 1597 TACCTCTTCCCAACAGGAGCTATTTATGCTAGCTGTGCAAGAACTGTTCAGTCCC 1656
Db 1964 -----AlaAlaProGlyAlaThrSerProSerAla-----SerSer 1975
QY 1657 TCTCTTCTCAATAAAGCATCTTCAA 1683
Db 1976 SerSerThrSerThrArgHisLeuGln 1984
RESULT 6
YHLL_EBV
ID YHLL_EBV STANDARD; PRT; 660 AA.
AC P03181;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical BHLPI protein.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84270667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrell B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984).
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
```

```
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; V01555; -: NOT_ANNOTATED_CDS.
DR PIR; A03742; Q0BE3.
KW Hypothetical protein; Early protein; Repeat.
FT DOMAIN 149 648 4 X 125 AA TANDEM REPEATS.
FT REPEAT 149 273 1.
FT REPEAT 274 398 2.
FT REPEAT 399 523 3.
FT REPEAT 524 648 4.
SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

Alignment Scores:
Pred. No.: 0.000466 Length: 660
Score: 192.00 Matches: 145
Percent Similarity: 32.28% Conservative: 49
Best Local Similarity: 24.13% Mismatches: 192
Query Match: 6.14% Indels: 215
DB: 1 Gaps: 36

US-09-762-021a-1 (1-1710) x YHLL_EBV (1-660)
QY 19 CAGCTCGCAGAAAGCTCTGGAGGAAGAGCTGGAGCAAAAGACCTCGACTTGGAGGCTTCA 78
Db 177 ArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThrProAla 196
QY 79 GCCAGGCCAGACAGATGGAGGCGCTGATGGAAGAGCGCTCCCTATGTAGGACAGGC 138
Db 197 Ala---ProGlyProGlyGlyAla-----AlaValProSerGlyAlaThr 211
QY 139 ACGCTA-----TCTGGAGCGGGGATCCCTCCAGAACAGCC-----CCACCA 180
Db 212 ProHisProGluArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuPro 231
QY 181 GAG-----GACCCCTAGA-----192
Db 232 GluArgGlnGluProArgLeuProGlnAspLeuAlaAlaGlnArgCysProAlaGly 251
QY 193 -----GCACAGCCTCCACCACATC 210
Db 252 ProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCys 271
QY 211 CCCAAGGCCCTGCCACGCCACAGTCCCGAGAACCAAGTGTCTTACTCTGCCTCC 270
Db 272 ProArgSerAlaArgAsnPro---GlyCysProArgThrTirpArgArgSerGlyAla 290
QY 271 TCCAAG-----GCCGTCTCTTCCCGAGGAGCCAGAGGAGGACGA 312
Db 291 GlnArgGlyHisProProGlyAlaGlyGlnArgProSerGlyProThrGlyArg 310
QY 313 GGAAGT-----GCTGAACCATGTCTCTAAGGAGCATTTGAGCT 348
Db 311 ProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAla 330
QY 349 GTT-----CATGGAAAGCTGGAAAGGCCAGGCCAGCAAGACAGCAG 390
Db 331 ValProSerGlyAlaThrProHisProGlnArgGlySerGlyProAlaAspProAla 350
QY 391 GAAGAAATTTGGGAAAAAACAAGCAGGAGGTCTCACCAGGACAGTACAT 450
Db 351 AlaAlaArgLeuProProGluArgGlnPro-----CCTCTGGG 361
QY 451 TGACTGCTTCAGAGATCAAGTACAGCTTCAA-----CCTCTGGG 492
Db 362 ---ArgLeuProGlnAspLeuAlaAlaGlnArgCysProAlaGlyProProProThr 380
QY 493 AAGGTGCCCTGGCTGAAGAGACAAAGTGCCTGAGCTCGTACAT-----CCT 546
Db 381 ArgSerGlyAlaAlaAlaGln-----ArgThrHisArgArgPro 393
QY 547 CTTCAGTCCCTGAA-----CTTCATCTGGCAGGTGCGCTGAGGTGG-----591
Db 547 CTTCAGTCCCTGAA-----CTTCATCTGGCAGGTGCGCTGAGGTGG-----591
```

Db 394 ProGlyCysProArgSerAlaArgAsnProGly---CysProArgThrTrpArgArgArg 412
Qy 592 -----CCTAGCAGCCCAAGTGATGATCATCCCTCTCACCCTCA 630
Db 413 SerGlyAlaGlnArgGlyHisProProProGlyAlaGlyGlnArgProSerGlyProThr 432
Qy 631 AGTATCAACCTGCTACAGCTCTGTCTAAGCCACCTGAGAGTAACCTTGGATGGGGTT 690
Db 433 GlyGlyArgProAla--AlaPro----- 439
Qy 691 GGGCCAGCCTGGGA---CCACTAGCC---GGCGGAGCTGGACAGGCGATGAGCCCTGGCC 744
Db 440 --GlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAlaValProSerG 459
Qy 745 CTACCAACCCACATCTCTCAGATGACTGGCAACTTCCAGAGCCCTCCAGCCCAAGCACCCCTT 804
Db 459 lYAlaThrProHisProGlu---ArgGlySerGlyProAlaAspPro-ProAlaAlaAla 477
Qy 805 AGGATACAGGACCTGTTCCCTTCGGCGGGGAAGTCATAGTTAGGAGCACCTTCACA 864
Db 478 ArgLeuPro-----ProGluArgGlnGluProArgLeuProGlnAspLeuAla 493
Qy 865 CTTTCCTCAGAGAGACACACACACCATGACCTCAGCCTGGGACCCCACTCCAGGCC 924
Db 494 AlaAlaGlnArgCysProAlaGlyProProProThrArgSerGlyAlaAlaAlaGlnArg 513
Qy 925 CTC-----CAGCCCCAAACCTGCCAGCAGCCCTGAAAATGCAAGTCTGTACGAGTT 978
Db 514 ThrHisArgProProGlyCysProArgSerAlaArgAsnProGlyCysProArgThr 533
Qy 979 TGAAGCTAGGAACCCAGGGAAGTACTGTGGT-----CCAGGGAGAGAAAGCTGGA 1029
Db 534 TrpArg---ArgArgSerGlyAlaGlnArgGlyHisProProGlyAlaGlyGlnArg 552
Qy 1030 GGTCTCGACACAGCAAGCGGTGGTGGCTGCTGCAAGAATGAGCGGAGCGGCGGCTA 1089
Db 553 ProSerGlyProThr-----GlyGlyAlaAla-----GlyGlyArgProAlaAla 563
Qy 1090 CATTCCAAGCAACATCTCGAGCCCTTACAGCGGGGACCCCTGGGACCCAGGCGCCAGTC 1149
Db 564 -----ProGlyAlaPro-----GlyThrProAlaAlaProGlyPro--- 575
Qy 1150 ACCCTCGGGTTCCTCAATGCTCGACTTACGAGCTCGAGGCGCTGAAGAGTACACAGCTGGCT 1209
Db 576 -----GlyGlyGlyAlaAla----- 580
Qy 1210 GCAGGCAGAGAACTTCTCCACTGCCAGCGTGGAGCACACTTGGGTCCCTGACGGGACCCA 1269
Db 581 -----ValProSerGly----- 584
Qy 1270 GCTACTTCGCATAAGACCTGGGAGCTPACAGATGCTATGTCCAGAGGAGCCCAACGAA 1329
Db 585 AlaThrProHisProGluArgGlySer-----GlyProAlaAsp 597
Qy 1330 CCGTCCCGCTGGAGCTGTCAAGAGGATGCTGGGGATAAGCCCTTAGCCACCACTTA 1389
Db 598 Pro-----ProAlaAlaAlaArgLeu 604
Qy 1390 GACACCTCCAGACAGCCCGCTGATGCAAGATGGCAGATCTGATACCCATTAGAGC 1449
Db 605 ProProGluArgGlnGluProArgLeu----- 613
Qy 1450 CCCGAGAAATCTCTCTTGGATCCAGTTTTCAGCAAAACCCCAACCCAGCTCACACAG 1509
Db 614 ProGlnAspLeuAlaAlaGlnArgCysProAlaGlyProProProThrArg----- 631
Qy 1510 CAAACAAATGGACAGCCCGAGGGCTGAA---GCAACAGTGTCCCTCTGGCTGT 1563
Db 632 -----SerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCys 646
RESULT 7
SHKL_RAT
ID SHKL_RAT STANDARD; PRT; 2167 AA.

AC Q9WV48; Q9WU13; Q9WUB8; Q9QZ28;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE SH3 and multiple ankyrin repeat domains protein 1 (Shank1) (CKAP/SAPAP
DE interacting protein) (SPANK-1) (Synamon) (Somatostatin receptor
DE interacting protein) (SSTR interacting protein) (SSTRIP).
GN SHANK1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=101116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND INTERACTION WITH DLGAP1 AND
RP DLG4.
RC TISSUE=Brain;
RX MEDLINE=99419021; PubMed=10488079;
RA Yao I., Hata Y., Hirao K., Deguchi M., Ide N., Takeuchi M., Takai Y.;
RT "Synamon, a novel neuronal protein interacting with synapse-associated
RT protein 90/postsynaptic density-95-associated protein.";
RL J. Biol. Chem. 274:27463-27466(1999).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 4), AND INTERACTION WITH DLGAP1.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=99360650; PubMed=10433268;
RA Naisbitt S., Kim E., Tu J.C., Xiao B., Sala C., Valtchanoff J.,
RA Weinberg R.J., Worley P.F., Sheng M.;
RT "Shank, a novel family of postsynaptic density proteins that binds to
RT the NMDA receptor/PSD-95/GKAP complex and cortactin.";
RL Neuron 23:569-582(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC TISSUE=Brain;
RX MEDLINE=20549637; PubMed=10958799;
RA Tobaben S., Suedhof T.C., Stahl B.;
RT "The G protein-coupled receptor CL1 interacts directly with proteins
RT of the shank family.";
RL J. Biol. Chem. 275:36204-36210(2000).
RN [4]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND
RP DEVELOPMENTAL STAGE.
RC TISSUE=Brain;
RX MEDLINE=99436166; PubMed=10506216;
RA Lim S., Naisbitt S., Yoon J., Hwang J.I., Suh P.G., Sheng M., Kim E.;
RT "Characterization of the Shank family of synaptic proteins. Multiple
RT genes, alternative splicing, and differential expression in brain and
RT development.";
RL J. Biol. Chem. 274:29510-29518(1999).
RN [5]
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 4).
RC TISSUE=Brain;
RX MEDLINE=20020275; PubMed=10551867;
RA Zitzer H., Hoenck H.-H., Baechner D., Richter D., Kreienkamp H.-J.;
RT "Somatostatin receptor interacting protein defines a novel family of
RT multidomain proteins present in human and rodent brain.";
RL J. Biol. Chem. 274:32997-33001(1999).
RN [6]
RP INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.
RX PubMed=10433269;
RA Tu J.C., Xiao B., Naisbitt S., Yuan J.P., Petralia R.S., Brakeman P.,
RA Doan A., Akalu V.K., Lanahan A.A., Sheng M., Worley P.F.;
RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of
RT postsynaptic density proteins.";
RL Neuron 23:583-592(1999).
RN [7]
RP INTERACTION WITH SPTAN1.
RX PubMed=11509555;
RA Bockers T.M., Mameza M.G., Kreutz M.R., Bockmann J., Weise C.,
RA Buck F., Richter D., Gundelfinger E.D., Kreienkamp H.-J.;
RT "Synaptic scaffolding proteins in rat brain. Ankyrin repeats of the
RT multidomain Shank protein family interact with the cytoskeletal
RT protein alpha-Fodrin.";
RL J. Biol. Chem. 276:40104-40112(2001).

Db 331 aAlaAlaAlaProAlaProGlyValLysAlaGluSerProLysArgValValGlnAl 351
 Qy 1005 CTGTGCTCCAGGAGAGAGCT---GGAGGTCTTGACACACAGCAAGCGGTGGTGG 1061
 Db 331 aAlaProProAlaAlaGlnThrLeuAlaAlaSerGlyProAlaSerThrAlaAlaSer-M 371
 Qy 1062 TGAAGATAGCGGAGCGGAGCGGTACATTCACAGC---AACATCTGGAGCCCTTAC 1118
 Db 371 etValIleGlyProThrMetGlnGlyAlaLeuProSerProAlaAlaValProProAla 391
 Qy 1119 AGCCGGGACCCCT---GGACCCAGGGCCAGTCACCT 1154
 Db 391 laProGlyThrProThrGlyLeuProLysGlyAlaAlaGlyAlaValThrGlnSerLeu 411
 Qy 1155 CTCGGGTTCACATG---CTTCGACTTAGCTCGAGCGCTCAAGAGGPCA 1199
 Db 411 erArgThrProThrAlaThrThrSerGlyIleArgAlaThrLeuThrProThrValLeu 431
 Qy 1200 CAGACTGGCTCAGCAGAGAACTTCACCTGCACGGTGAGGACACTTGGTCCCTGA 1259
 Db 431 laProArgLeuProGlnProProGlnAsnProThrAsnIleGlnAsnPheGlnLeuProp 451
 Qy 1260 CGGGAGCCACTACTTCGCATAGACCTGGGAGCTACAGATGCTATGTCACAGGAGG 1319
 Db 451 roGlyMetValLeuValArgSerGluAsnGlyGlnLeuLeuMetIle---ProGlnGln 470
 Qy 1320 CCCACGAATCTGTCTCCGGCTGAGGCTGCAGAGGACTGCTGGGATAAGCCCTTAGG 1379
 Db 470 la-----LeuAlaGlnMetGlnAlaGlnAlaHisAlaGlnProGlnThrThrMet 487
 Qy 1380 CACCAGCTTAGACACTCCAGAACACAGGCGCCGCTGATGCAAGATGSCAGATCTGATAC 1439
 Db 487 laProArgProAlaThrProThrSerAlaPro-----ProValGlnIleSer 503
 Qy 1440 CCATTAGACCCCGAGATTCCTCTCTGGATCCAGTTTCGACCAACCCACACCCCA 1499
 Db 503 hrValGlnAlaProGlyThrProIleIle-----AlaArgGlnValThr-----Prot 519
 Qy 1500 GCTCACAGCAAAACAAATGGACAGGCGCCAGAGGCTGAACCAACAGTGTCCCTCTG 1558
 Db 519 hrThrIleIleGlnValSerGlnAlaGlnThrThrValGlnProSerAlaThrLeu 538
 RESULT 9
 YQ36_CAEEL STANDARD; PRT; 963 AA.
 AC Q09457;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Putative cuticle collagen C09G5.6.
 GN C09G5.6.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID:6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-Bristol N2;
 RA Palmer S.;
 RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: NEMATODE CUTICLES ARE COMPOSED LARGELY OF COLLAGEN-LIKE
 CC PROTEINS. THE CUTICLE FUNCTIONS BOTH AS AN EXOSKELETON AND AS A
 CC BARRIER TO PROTECT THE WORM FROM ITS ENVIRONMENT (BY SIMILARITY).
 CC -!- SUBUNIT: COLLAGEN POLYPEPTIDE CHAINS ARE COMPLEXED WITHIN THE
 CC CUTICLE BY DISULFIDE BONDS AND OTHER TYPES OF COVALENT CROSS-
 CC LINKS (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE CUTICULAR COLLAGEN FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; Z46791; CAA86755.1; -;
 DR WormPep; C09G5.6; CE01486.
 DR InterPro; IPR002486; Col_cuticle_N.
 DR InterPro; IPR000087; Collagen.
 DR Pfam; PF01391; Collagen; 3.
 DR Pfam; PF01484; Col_cuticle_N; 1.
 KW Hypothetical protein; Cuticle; Connective tissue; Repeat;
 KW Multigene family; Collagen.
 FT DOMAIN 392 423 TRIPLE-HELICAL REGION.
 FT DOMAIN 441 503 TRIPLE-HELICAL REGION.
 FT DOMAIN 506 567 TRIPLE-HELICAL REGION.
 FT DOMAIN 663 666 POLY-PRO.
 FT DOMAIN 685 688 POLY-PRO.
 SQ SEQUENCE 963 AA; 107031 MW; AFF895A75909F66E CRC64;
 Alignment Scores:
 Pred. No.: 0.00117 Length: 963
 Score: 185.00 Matches: 141
 Percent Similarity: 30.94% Conservative: 44
 Best Local Similarity: 23.58% Mismatches: 206
 Query Match: 5.92% Indels: 208
 DB: 1 Gaps: 31
 US-09-762-021a-1 (1-1710) x YQ36_CAEEL (1-963)
 Qy 34 TCTGGAGAGAGCT-----GGAGCAAGACCTCGACTTGGAGGCGCT----- 75
 Db 154 SerGlyPheArgAlaProProAlaAlaThrSerThrTyArgProProHisGlySer 173
 Qy 76 -----TCAGCCAGG-----TCAGCCAGG-----CCAGGACAG 93
 Db 174 AsnTyrraspasntyrGlyArgGluProAlaSerSerArgArgProTyrrProGlnGln 193
 Qy 94 ATGGAGGGGGCTCTATGGAAGCGCGCTCCCTATGGAGCAGGACGCTATCTGGAGCC 153
 Db 194 ProProSerThrSerAlaProHisSerSerProAsnAsnArgThrSerLeuTyrrAsnPro 213
 Qy 154 GGGATCCCTCCAGAACAGCCCGCCAGGAGGACCCCTAGAGCA-----CAG 198
 Db 214 GlnProProLysThrGlyTyrrProThrAsnProArgValProTyrrAsnProGln 233
 Qy 199 CCT-----CCACCATCCCT-----AAGGCCCTGCCACGCCACAC 234
 Db 234 ProAsnTyrrThrArgGlnProThrTyrrProGluaspasnArgAlaProTyrrLysProThr 253
 Qy 235 CAGTCCCGCAGAACCAAGTGCCTTTACTCTGCCTCTCTCAAGCGGCTCTCTTCCCGCGA 294
 Db 254 ArgSerProAsnThrProPro-----ProArg 262
 Qy 295 GGACCCAGAGAGGAGGAGGAGGAGTGTGAACCATGTCTTAAGGACATTGAGCTGTTCAT 354
 Db 263 GlnProSerGlyTyrraspSerAspGlyGlnThrProProSerSerProArgIleTyrr 282
 Qy 355 GGGAAAGCTGGAGAGAGGCCCGCCAGGACCAAGACCCAGCAGGAAGAAATTTGGGAAAAAAA 414
 Db 283 -----AsnThr 284
 Qy 415 CAAGACACAGGAGGTCTCACCCAGGACGACAGTACATTGACTTCCAGAGATCAAGTA 474
 Db 285 ArgArgProAsnAsnHisGlyProGly-----TyrProGluaspGlnVal 299
 Qy 475 CAGCTTCAACCTCTCGGAAGGCTGGCCACCTGGCTGAAGGACAGACAGTCCCTGAGCT 534
 Db 300 ProThrAlaProPro-----ValProGlyGlnGln----- 309
 Qy 535 CGTACACATCTCTTCAAGTCCCTGAACCTTCATCTCCTGCCAGGTGCCCTGAGGCTGCGCT 594
 Db 310 ArgValProProGlnThrArgAsn---ProProAsnProThrAsnThrArgGlnPro 328
 Qy 595 AGCAGCCCAAGTGATCTCACCCCTCTCACCCCTCTCACCCCTAAAAGCTATCAACCTGTACAGTCTGT 654

DR	InterPro:	IPR000087;	Collagen.
DR	InterPro:	IPR001007;	VWF_C.
DR	Pfam:	PF01391;	Collagen; 17.
DR	ProDom:	PD000007;	Collagen; 1.
DR	PROSITE:	PS01208;	WMFC; PARTIAL.
KW	Extracellular matrix;	Connective tissue;	Repeat; Hydroxylation;
KW	Protein domain;	Collagen.	
FT	DOMAIN	1	14
FT	DOMAIN	15	1040
FT	DOMAIN	1041	1049
FT	MOD_RES	95	95
FT	MOD_RES	107	107
FT	MOD_RES	119	119
FT	MOD_RES	938	938
FT	MOD_RES	950	950
FT	CARBOHYD	107	107
FT	CARBOHYD	950	950
FT	DISULFID	1040	1040
FT	DISULFID	1041	1041
SO	SEQUENCE	1049 AA;	93651 MW; 8EPC33DI0C66EC9A3 CRC64;

Alignment Scores:	
Pred. No.:	0.00125
Score:	184.50
Percent Similarity:	29.92%
Best Local Similarity:	24.72%
Query Match:	5.90%
DB:	1

US-09-762-021A-1 (1-1710) x CAL3_BOVIN (1-1049)	
Qy	10 ACTGAAGACCAGCCTCGAGAA---GGCTCTGGAGGAGAGACTGGACGAAGACCTCGACT 66
Dy	490 ThrSerGlyProProGlyGluAsnGlyLysProGlyGluProGlyLysGlyGluAla 509
Qy	67 TGGAGGCCTTCAGCCAGGCCAGACAGATGGAGGGGCGTGTATGGAAAGCGCGTCCC 126
Dy	510 GlyAlaProGlyIleProGlyLysGlyAspSerGlyAlaProGlyGluArgGlyPro 529
Qy	127 TATGGACA-----GGCAGCTATCTGGAGCGGGGATCCCTCCAGACAGCC 174
Dy	530 ProGlyAlaGlyLysProProGlyProArgGlyGlyAlaGlyProProGly----- 546
Qy	175 CCACGAGGACCCCTAGA-----GCACAGCCTCCCAACCATCCCCAAGGCCCTGCCAG 228
Dy	547 --ProGluGlyLysGlyAlaAlaGlyProProGlyProGlySerAlaGlyThr 565
Qy	229 CCACACCAAGTCCGGAAACCAAGTCGCTTTACTCTGCTTCCCAAGGCGGTCTCTTTC 288
Dy	566 Progly-----Leu 568
Qy	289 CCCCAGGAGCCAGAGAGGACGAGNAAGTGCTGAACCATGTCTTAAGGCATTTGACGT 348
Dy	569 GlnGlyMetProGlyGluArGlyGlyProGlyProGlyProGlyProGlyAspLysGly 588
Qy	349 GTTCATGGGAAAGCTGA-----GAAGGCCAGGCAAAGACACGAGGAAGAAGAAATT 402
Dy	589 GluProGlySerSerGlyValaspGlyAlaProGlyLysasp----- 602
Qy	403 TGGGAAAAAACAGGACCA---GGAGGTCTCACCCAGGCACAGTACATTGACTGCTT 459
Dy	603 -----GlyProArgGlyProThrGlyPro----- 610
Qy	460 CCAGAAGATCAAGTACAGCTTCAACCTCTCTGGGAAGGCTGGCCACCTGGCTGAAGGAC 519
Dy	611 -----IleGlyProProGlyProAlaGlycIn-----ProGlyAsp 622
Qy	520 AAGTGCCCTTGAGCTCGTACACATCTCTTCAAGTCCCTGAACTTCATCTCGCCAGGTG 579
Dy	623 Lys-----GlyGluSerGlyAlaProGlyValProGlyIle-----AlaGly 636
Qy	580 CCCTGAGGCTGGCTAGCAGCCCAAGTGTCTCACCCCTCTCACCCCTATAAGCTATCAA 639

```
QY 1498 CAGCTCACACAGCAAAACAATGGACAGGCCAGAGGCTGAAGCAAAACAGTGTCTC----- 1551
D 966 -----GlyProAlaGlyHisGlnGlyAlaValGlySer 976
QY 1552 -----CCTTCTGGCTGTGTGGAGGCTCC 1575
D 977 ProGlyProAlaGlyProArgGlyProValGlyProSer 989

RESULT 11
DRPL_RAT
ID DRPL_RAT STANDARD: PRT: 1183 AA.
AC P54258;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Atrophin-1 (Dentatorubral-pallidoluysian atrophy protein).
GN DRPLA.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Cerebellum, and Striatum;
RX MEDLINE=97317138; PubMed=9173996;
RA Loev S.J., Margolis R.L., Young W.S., Li S.-H., Schilling G.,
RA Ashworth R.G., Ross C.A.;
RT "Cloning and expression of the rat atrophin-1 (DRPLA disease gene)
RT homologue.";
RL Neurobiol. Dis. 2:129-138(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain, Cerebellum, Hippocampus, and Substantia nigra;
RX MEDLINE=96081227; PubMed=8541849;
RA Schmitt I., Epplen J.T., Riess O.;
RT "Predominant neuronal expression of the gene responsible for
RT dentatorubral-pallidoluysian atrophy (DRPLA) in rat.";
RL Hum. Mol. Genet. 4:1619-1624(1995).
CC -I- TISSUE SPECIFICITY: PREDOMINANT NEURONAL EXPRESSION, ALTHOUGH
CC MARKEDLY REDUCED AMOUNTS ARE FOUND IN MOST OTHER TISSUES.
CC -I- DEVELOPMENTAL STAGE: SIMILAR EXPRESSION AT ALL DEVELOPMENT STAGES
CC (DAY 14.5 P.C., 17.5 P.C., NEWBORNS AND ADULTS).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U31777; AAA80337.1; -
DR EMBL: X89453; CAA61623.1; -
DR InterPro: IP0002951; Atrophin.
DR Pfam: PF03154; Atrophin-1; 2.
DR PRINTS: PR01222; ATROPHIN.
FT DOMAIN 165 171 POLY-PRO.
FT DOMAIN 303 306 POLY-PRO.
FT DOMAIN 377 383 POLY-SER.
FT DOMAIN 387 391 POLY-SER.
FT DOMAIN 440 446 POLY-PRO.
FT DOMAIN 477 480 POLY-HIS.
FT DOMAIN 481 489 POLY-GLN.
FT DOMAIN 502 505 POLY-PRO.
FT DOMAIN 562 572 POLY-SER.
FT DOMAIN 702 705 POLY-PRO.
FT DOMAIN 455 455 N -> S (IN REF. 2).
FT CONFLICT 594 594 F -> L (IN REF. 2).
FT CONFLICT 689 689 P -> R (IN REF. 2).
FT CONFLICT 717 717 T -> M (IN REF. 2).
FT CONFLICT 737 737 A -> V (IN REF. 2).
FT CONFLICT 965 965 MISSING (IN REF. 2).
```

```
SQ SEQUENCE 1183 AA; 124778 MW; 7FB9928DCA9F9B1F CRC64;
Alignment Scores:
Pred. No.: 0.00125 Length: 1183
Score: 184.50 Matches: 154
Percent Similarity: 32.32% Conservative: 49
Best Local Similarity: 24.52% Mismatches: 196
Query Match: 5.90% Indels: 229
DB: 1 Gaps: 34

US-09-762-021a-1 (1-1710) x DRPL_RAT (1-1183)
QY 34 TCTGGAGGAAGAGCTGGAGCAAAACACCTCGACTTGGAGGCTTCAGCAGGCCAGGACGAGCAG 93
D 265 SerSerGlyAlaSerGlyAla-----ProProAlaLysProProAsn 278
QY 94 ATGGAGGGGGCTGCTATGTGGAAGGCCCTCCCTATATGGAGGAGGACGCTATCTCGAGGCC 153
D 279 -----ThrProValGlyAlaGlyAsnLeuProSerAla 289
QY 154 GGGATCCCTCCAGAAACAGCCCAAGAGGACCTAGAGCAGACAGCTCCCAACCATCCCC 213
D 290 ProProProAla-----ThrPheProHisValThrProAsnLeuProPro-ProProAlaLe 308
QY 214 AAGGCCCTGCCACGCCACACAGTGGCCGAGAACCAAGTGCCTTTACTCTGCTCTCTCC 273
D 308 uArgProLeu---AsnAsnAlaSerAlaSerProProGlyMetGlyAlaGlnProIlePr 327
QY 274 AAGGC-----GGTCTCTTCCCGGAGGA 297
D 327 o-GlyHisLeuProSerProHisAlaMetGlyGlnGlyMetSerGlyLeuProProGlyP 347
QY 298 CCCAGAGGAGGAGGAGGAGGAGTCTCTCAACCATGCTCTTAAGGGACATTTAGCTGTTTCATGG 357
D 347 roGluLysGlyPro-----ThrLeuAlaProSer----- 356
QY 358 AAGCTGGAGGAAGGCCCGAGCAAGACAGCAGGAGGAAGAAATTTGGGAAAAAACA 417
D 357 -----ProHisProLeuProProAlaSerSerSerAlaProGlyProPro 372
QY 418 GCACCGAGGAGGTCTCACCCAGGCACATGACATTCACCTTCCAGAGAGATCAAGTACAG 477
D 372 eArgTyProTyProSerSerSerSerSerSerSerValAlaAlaSerSerSerSerA 392
QY 478 CTTCACCTCTCTGGGAGGCTGGCCACCTGGCTGAAGGAGCAAGTGCC----- 527
D 392 leAlaThrSer-----GlnTyProAlaSerGlnT 402
QY 528 --CTGAGCTGTGTACATCTCTTCAAGTCCCTGAACTTCATCTCGGCGCAGTGCCTGA 585
D 402 hrLeuProSerTyProHisSerPheProPro-----ProThrSerMets 417
QY 586 GGCTGGCTAGCAG---CCCAGTGTATCTACCCCTCTCACCCTCTCACCCTAAGCTATCAACCT 642
D 417 erValSerAsnGlnProProLysTyThrGlnProSerLeuProSerGlnAlaVal----- 435
QY 643 GCTACAGCTCTCTTAAGCCACCTGAGAGTAACCTTTTGGATGGGTGGGCCAG----- 698
D 436 -----TrpSerGlnGlyP 440
QY 699 -----CCTGGAC 705
D 440 roProProProProProTyGlyArgLeuLeuProAsnAsnAsnThrHisProGlyP 460
QY 706 CACTAGCGGGCGGACTGA-----CAGGCCATGAGCCCTGCTTACCAACCCAC 756
D 460 roPhePro---ProThrGlyGlyGlnSerThrAlaHisProProAlaHisHisH 479
QY 757 ATTCTCAGATGACTGGCAACTTCCAGAGCCCTCCAGCCAA----- 796
D 479 isHisGlnGlnGln--GlnGlnProGlnProGlnProGlnGlnHisHisG 498
QY 797 -----GCACCTTAGGATACAGGACCTGTTTCCCTTCGCGGGGGA 840
```

```

Db 498 yAsnSerGlyProProGlyAlaThrProHisProLeuGlu----- 513
QY 841 TCATAGTTAGGAGCACCCTCAGCTCTTCTCAGGAGAAGACACAAACCATGACCTCA 900
Db 514 -----SerSerAsnSerHis-----HisAlaHisProTyrAsnMetSerPro-S 528
QY 901 GCGTGGGACCAACTCCAGCCCTCCAGCCCA---AACCTGCCAGCCAGCCCTGAA 957
Db 528 erLeuGlySerLeuArgProTyrProGlyProAlaHisLeuPro----- 544
QY 958 ATGCAAGTCTGTAGGAGTTTGAAGCTAGGAAACCCAGGAGCTGCTGGTCCAGGG 1017
Db 545 -----SerHisGlyGln-----ValSerTyrS 552
QY 1018 AGAAGAGCTGGAGTTCTGGACACAGCAAGCGGTGGTGGTGAAGATGATGCGGG 1077
Db 552 er-GlnAlaGlyProAsnGlyProProValSerSerSerSerSerSerSerSer 571
QY 1078 ACGAGCGGCTACATTCCAAGCAACATCCT-----GGAGCCCTACAGCCGG 1125
Db 572 SerGlnAlaAlaTyrSerCysSerHisProSerSerSerGlnGlyProGlnGlyAlaSer 591
QY 1126 GACCCCTGGGACCCA---GGGCCATGACCTCTCGGTTCCAATGCTTCGACTAGCTC 1182
Db 592 TyrProPheProProValProProIleThrThrSerSerAlaThrLeuSerThrValIle 611
QY 1183 -----GAGCCTGAAGAGTGCACAGACTGCTGCAGGAGGAGCACTTCTCCAC 1230
Db 612 AlaThrValAlaSerSerProAlaGlyTyrIleThrAlaSer----- 625
QY 1231 TCCACGCTGAGGACACTTGGTCCCTGACGGGGAGCCA----- 1269
Db 626 -----ProGlyProProGlnTyrSerLysArgAlaPro 637
QY 1270 ---GCTACTTCCATAAGACCTGG-----GGAGCTACAGATGCT 1305
Db 638 SerProGlySerTyrIleThrAlaThrProGlyTyrLysProGlySerProSer 657
QY 1306 ATGTCACAGGAGGCC---ACGAATCCTGTCCCGCTGGAGGCTGT 1350
Db 658 pheArgThrGlyThrProProGlyTyrArgGlyThrSerPro---ProAlaGlyProGly 676
QY 1351 CAGAAGATGCTGGGATAAGCCCTTAGCACCAGCTTAGACACCTCCAAGAACAGGCC 1410
Db 677 ThrPheLysProGlySerProThrValGly-----ProGly 688
QY 1411 CCGCTGATGCAAGATGGCAGATCTGATCCCATAGACCCCGGAGAAATCTCTCTGGA 1470
Db 689 ProLeuProProAlaGly-----ProSerSerLeuSerSerLeu 701
QY 1471 TCCAGTTTGCAGCAAAACCCACACCCAGCTCACAGCAAAACAATGGACAGGCCA 1530
Db 702 ProProProProAlaAlaProThr-----ThrGlyPro 712
QY 1531 GAGGCTGAAGCAACAGTGTCCCTCTGCTGTGTGGAGCTCCCGACAGTACACCTAT 1590
Db 713 ProLeuThrAlaThrGlnIleLysGln-----GluProAlaGluGluTyr 727
QY 1591 TTATTTTACCTTTTCCCAACCTGGAGCATTTATGCTTAGCTTGCTCAAGAACTGTC 1650
Db 728 GluThrProGluSerProValProProAlaArgSerPro----- 740
QY 1651 AGTCCTCTCTCT 1662
Db 741 SerProProPro 744
RESULT 12
CA17_HUMAN
ID CA17 HUMAN STANDARD; PRT; 2944 AA.
AC Q02388; Q14054; Q16507;
DT 01-JUN-1994 (Rel. 23, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)

```

```

DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Collagen alpha 1(VII) chain precursor (Long-chain collagen) (LC
GN COL7A1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eumalia; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94327588; PubMed=8051117;
RA Christiano A.M., Greenspan D.S., Lee S., Uitto J.;
RT "Cloning of human type VII collagen. Complete primary sequence of the
RT alpha 1(VII) chain and identification of intragenic polymorphisms.";
RL J. Biol. Chem. 269:20256-20262(1994).
RN [2]
RP SEQUENCE OF 128-1493 FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=93338437; PubMed=1307247;
RA Christiano A.M., Rosenbaum L.M., Chung-Honet L.C., Parente M.G.,
RA Woodley D.T., Pan T.C., Zhang R.Z., Chu M.-L., Burgeson R.E.,
RA Uitto J.;
RT "The large non-collagenous domain (NC-1) of type VII collagen is
RT amino-terminal and chimeric. Homology to cartilage matrix protein,
RT the type III domains of fibronectin and the A domains of von
RT Willebrand factor.";
RL Hum. Mol. Genet. 1:475-481(1992).
RN [3]
RP SEQUENCE OF 815-1439 FROM N.A.
RX MEDLINE=91334380; PubMed=1871109;
RA Parente M.G., Chung L.C., Ryyanen J., Woodley D.T., Wynn K.W.,
RA Bauer E.A., Mattei M.-G., Chu M.-L., Uitto J.;
RT "Human type VII collagen: cDNA cloning and chromosomal mapping of the
RT gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 88:6931-6935(1991).
RN [4]
RP SEQUENCE OF 369-1255 FROM N.A.
RX MEDLINE=93107742; PubMed=1469284;
RA Gammon W.R., Abernethy M.L., Padilla K.M., Prisayan P.S.,
RA Cook M.E., Wright J., Briggman R.A., Hunt S.W. III;
RT "Noncollagenous (NC1) domain of collagen VII resembles multidomain
RT adhesion proteins involved in tissue-specific organization of
RT extracellular matrix.";
RL J. Invest. Dermatol. 99:691-696(1992).
RN [5]
RP SEQUENCE OF 340-675 FROM N.A.
RC TISSUE-Keratinocytes;
RX MEDLINE=92231902; PubMed=1567409;
RA Tanaka T., Takahashi K., Furukawa F., Imamura S.;
RT "Molecular cloning and characterization of type VII collagen cDNA.";
RL Biochem. Biophys. Res. Commun. 183:958-963(1992).
RN [6]
RP SEQUENCE OF 2395-2944 FROM N.A.
RX MEDLINE=93271985; PubMed=8499916;
RA Greenspan D.S.;
RT "The carboxyl-terminal half of type VII collagen, including the non-
RT collagenous NC-2 domain and intron/exon organization of the
RT corresponding region of the COL7A1 gene.";
RL Hum. Mol. Genet. 2:273-278(1993).
RN [7]
RP SEQUENCE OF 1-87 FROM N.A.
RC TISSUE-Placenta;
RX MEDLINE=94375010; PubMed=8088784;
RA Christiano A.M., Hoffman G.G., Chung-Honet L.C., Lee S., Cheng W.,
RA Uitto J., Greenspan D.S.;
RT "Structural organization of the human type VII collagen gene (COL7A1),
RT composed of more exons than any previously characterized gene.";
RL Genomics 21:169-179(1994).
RN [8]
RP REVIEW ON DEB VARIANTS.
RX MEDLINE=98041696; PubMed=9375848;
RA Jaervikallio A., Pulkkinen L., Uitto J.;
RT "Molecular basis of dystrophic epidermolysis bullosa: mutations in
RT the type VII collagen gene (COL7A1).";

```



```
QY 175 CCACGAGAGCCCTAGACACAGCCTCCACCATCCCAAGGCC---CCTGCCACGCCA 231
DB 1333 ProArgGlyAspProGlyGluArgGlyProArgGlyProGlyGluProGlyAlaPro 1352
QY 232 CACAGTCCCGAGAACCAAGTGCCTTACTCTGCTCTCCAGGCGGTCTCTCCGCC 291
DB 1353 GlyGlnValIleGlyGlyGluGlyProGlyLeuProGlyArgLysGlyAspProGlyPro 1372
QY 292 CGAGGACCCAGAGAGGAGGAGGAGT---CCTGAACCATGCTCTAAGGGACATTGAGCT 348
DB 1373 SerGlyProProGlyProArgGlyProLeuGlyAspProGlyProArgGlyProProGly 1392
QY 349 GTTCATGGAAAGTGGAGAGGCCAGGCAAGAACACAGCAGGAGGAAGAAATTGGGAA 408
DB 1393 LeuProGlyThrAlaMetLysGlyAspLysGlyAspArgGlyGlu-----Arg 1408
QY 409 AAAAAACAAGACAGGAGGTCTCAC-----435
DB 1409 GlyProProGlyProGlyGluGlyGlyIleAlaProGlyGluProGlyLeuProGlyLeu 1428
QY 436 CAGGCACAGTACATTGACTGCTTCCAGAAGATCAAGTACAGCTTCAACCTCTCTGGGAAG 495
DB 1429 ProGlySer-----ProGlyProGlnGlyProValGlyProProGlyLys 1443
QY 496 GCTGCCACCTGGCTGAAGGAGACAAGTGCCTGAGCTCGTACACATCTCTTCAAGTC 555
DB 1444 LysGlyGluLysGlyAspSerGluAsp-----GlyAla 1454
QY 556 CTGNACTTCATCTGGCCAGGTCGCTGAGGCTGAGGCTGAGCCGACGCCCAAGTGATCTCAC 615
DB 1455 ProGlyLeu---ProGlyGln-----ProGlySerProGlyGluGlnGly 1468
QY 616 CTCTCTCACCCCTAAAGCTATCAACCTGCTACAGCTCTGCTAAGCCACCTGAGAGTAA 675
DB 1469 ProArgGlyPro-----ProGly-----1474
QY 676 CCTTTGGATGGGTTGGCCCGAG-----CCTGGACCACTA-----GCC 713
DB 1475 -----AlaIleGlyProLysGlyAspArgGlyPheProGlyProLeuGlyGluAla 1491
QY 714 GGGCGGACTGACAGGCGATGAGCCCTGCTACCAACCCACATCTCAGATCACTGGC 773
DB 1492 GlyGluLysGlyGluArgGlyProProGlyProAlaGlySerArgGly---LeuProGly 1510
QY 774 AACTTCCAGACCCCTCCAGGCAACACCCCTTAGGATACAGGAGCTGTTTCCCTCGGCG 833
DB 1511 ValAlaGlyArgProGlyAlaLysGlyPro-Glu-----GlyProProGlyProThrG 1528
QY 834 GGGGAAGTCATAGTTAGGGAGCACCTCACACTTTCCTCAGGAGAAGACACACAACCA-- 891
DB 1528 yArgGlnGly-----GluLysGlyGluProGlyArgProG 1540
QY 892 -TGACCTCACCTGGGACCCCAACTC-----CAGGCCCTCAGGCCCAAACTG 941
DB 1540 yAspProAlaValGlyProAlaValAlaGlyProLysGlyGlyAspValG 1560
QY 942 CCCAGCCAGCCCTGAATGCTGTGTACAGAGTTTGAAGCTAGGAACCCAGGGAAC 1001
DB 1560 yProAlaGlyProArgGlyAlaThrGlyValGln-----1571
QY 1002 TGACTGTGGTCCAGGAGAGAAGCTGGAGGTTCTGGACCAAGCAAGCGGTGGGTGG 1061
DB 1572 -----GlyGluArgGlyProProGlyLeuValLeuProG 1583
QY 1062 TGAAGATGAGGGGCGGAGCGGCTACATTCCAGCAACATCTGGAGGCCCTACAGC 1121
DB 1583 yAsp-----ProGly-----1588
QY 1122 CGGGACCCCTGGACCCAGGCGGAGTCAACCTCTCGGGTTTCCAATGCTTCGACTAGCT 1181
DB 1588 sGlyAspProGlyAspArgGlyProIleGlyLeuThrGlyArgAlaGlyProProGlyAs 1608
QY 1182 CGAGGCTGAAGAGGTCACAGACTGGCTGCAGGCAGAGAA-----CTTCTCCA 1229
```

```
DB 1608 pSerGlyProProGlyGluLysGlyAspProGlyArgProGlyProGlyProValG 1628
QY 1230 CTGCCAGCGTGAGCACACTGGGTC-----CCTGACGGGGAGCC 1268
DB 1628 yProArgGlyArgAspGlyGluValGlyGluLysGlyAspGlyProProGlyAspPr 1648
QY 1269 AGCTACTTCGATTAAGACCTGGGAGCTACAGATGCTATGCTCCACAGGAGGCCACGAA 1328
DB 1648 oGlyLeuProGlyLysAla---GlyGluArgGlyLeuArgGlyAlaProGlyValArgG 1667
QY 1329 TCTGTCTCCCGCTGAGGCTGTCAGAAGGATGCTGGGATAAGCCCTTAGGCACACAGCTT 1388
DB 1667 yProValGlyGluLysGlyAspGlnGlyAspProGly-----1679
QY 1389 AGACACCTCCAGAACACAGGCCCGCTGATGTCAGATGCGCAGATCTGTATCCCATTTAG 1448
DB 1680 -----GluAspGlyArgAsnGly-----Se 1686
QY 1449 CCCCAGGAATTCC-----TCTTCTGGATCCAGTTTCAGCAAAACCCACACCC 1496
DB 1686 rProGlySerSerGlyProLysGlyAspArgGlyGluProGlyProGlyProGlyProG 1706
QY 1497 CCAGCTCACACAGCAAAACAATGGACAGGCCCGGAGGCTGAAGCAACAGTGTCCCT 1554
DB 1706 yArgLeuValAsp-----ThrGlyProGlyAlaArgGluLysGlyGluPro 1721

RESULT 13
MAPA_RAT
ID MAPA_RAT STANDARD; PRT; 2774 AA.
AC P34926;
AT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Microtubule-associated protein 1A (MAP 1A) [Contains: MAP1 light chain
DE LC2].
GN MAP1A.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=92355629; PubMed=1379599;
RA Langkopf A., Hammarback J.A., Mueller R., Vallee R.B., Garner C.C.;
RT "Microtubule-associated proteins 1A and LC2. Two proteins encoded in
RL J. Biol. Chem. 267:16561-16566(1992).
CC -|- FUNCTION: Structural protein involved in the filamentous cross-
CC bridging between microtubules and other skeletal elements.
CC -|- SUBUNIT: 3 different light chains, LC1, LC2 and LC3, can associate
CC with MAP1A and MAP1B proteins.
CC -|- TISSUE SPECIFICITY: BRAIN, HEART AND MUSCLE.
CC -|- DEVELOPMENTAL STAGE: EXPRESSED LATE DURING NEURONAL DEVELOPMENT
CC APPEARING WHEN AXONS AND DENDRITES BEGIN TO SOLIDIFY AND STABILIZE
CC THEIR MORPHOLOGY.
CC -|- DOMAIN: THE BASIC REGION CONTAINING THE REPEATS MAY BE RESPONSIBLE
CC FOR THE BINDING OF MAP1A TO MICROTUBULES.
CC -|- PTM: VARIOUS SERINE RESIDUES MAY BE PHOSPHORYLATED BY CAMP KINASE.
CC -|- PFM: LC2 IS CORRELATED WITH MAP1A. IT IS A POLYPEPTIDE GENERATED
CC FROM MAP1A BY PROTEOLYTIC PROCESSING. IT IS FREE TO ASSOCIATE WITH
CC BOTH MAP1A AND MAP1B.
CC -|- SIMILARITY: TO MAP1B.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```


Percent Similarity:	30.36%	Conservative:	30
Best Local Similarity:	25.00%	Mismatches:	201
Query Match:	5.79%	Indels:	189
DB:	1	Gaps:	31
US-09-762-021A-1 (1-1710) x CALL_CHICK (1-1453)			
QY	25	GCAGAGGCTCTGGAGAAAGCTGGAGCAAGAGCTCGACTTGG-----	69
DB	742	AlaAspGlyAlaProGlyLysAspGlyLeuArgGlyLeuThrGlyProIleGlyProPro 761	
QY	70	AGCCTTCACGCCAGGCCAGACAGATG-----GAGGGGCTGTATGGAAGGCCGCTCC 126	
DB	762	GlyProAlaGlyAlaProGlyLysGlyGlyAlaGlyProProGlyProAlaGlyPro 781	
QY	127	TATGGAGCA-----GGCAGCTATCT 147	
DB	782	ThrGlyAlaArgGlyAlaProGlyAspArgGlyGluProGlyProGlyProAlaGly 801	
QY	148	GGAGCGGGGATCCCTCCAGAACAGCCCA-----GAGGACCT 189	
DB	802	PheAlaGlyProGlyAlaAspGlyGlnProGlyAlaLysGlyLeuThrGlyAspAla 821	
QY	190	AGACACAGCCTCCCAACCATCCCAAGGCCCTGCCACGCCACACAGTGGCCGAGAAC 249	
DB	822	GlyAlaLysGlyAspAlaGlyProGlyProAlaGlyProThrGlyAlaPro----- 839	
QY	250	AAGTGCCTTACTCTCCCTCCCAAGGGGCTCTTCCCGAGGACCCAGAGGGA 309	
DB	840	-----GlyProAlaGly***ValGlyAlaPro---GlyProLysGlyAla 853	
QY	310	CGAGGAAGTCTGAACCATCTCTTAGGGACATTTGAGCTGTTTCATGGGAAAGCTGGAGAA 369	
DB	854	ArgGlySerAlaGly-----ProProGlyAlaThrGlyPheProGlyAlaAlaGly--- 870	
QY	370	GGCCAGGCAAGACAGCAGGAGGAAGAAATTTGGGAAAAAACAAGGACCGAGGAGG 429	
DB	871	-----ArgValGlyPro----- 874	
QY	430	TCTCACCCAGGCACAGTACATTGACTGCTTCCAGAGATCAAGTACAGCTTCAA---CCT 486	
DB	875	-----ProGly-----ProSerGlyAsnIleGlyLeuProGlyPro 886	
QY	487	CCTGGAGGCTGGCCACCTGCTGAAGGAGACAAGTGCCTGAGCTGTACATCCT 546	
DB	887	ProGlyProAlaGly-----Lys***GlySerLysGlyProArgGlyGluThrGlyPro 904	
QY	547	CTTCAAGTCCCTGAAGTCTATCTCTGGCCAGTGCCTGAGGCTGGCTAGGACCCAGT 606	
DB	905	Ala-----GlyArgProGlyGluProGly 912	
QY	607	GATCTCACCCCTCCACCCCTAAAGCTATCAACCTGCTACAGTCTGTCTAAGCCACC 666	
DB	913	ProAlaGlyProProGlyPro----- 919	
QY	667	TGAGAGTAACTTTGGATGGGTGGGCCCCAGCTGGACCACTAGCCGGGCGACTGGAC 726	
DB	920	-----ProGlyGlyLysGlySerProGlyAlaAspGlyProIleGly----- 933	
QY	727	AGGCGATGAGCCCTCCCTACCAACCCACATTCAGATGACTGGCAACTTCCAGAGCC 786	
DB	934	-----AlaProGlyThrProGly-----ProGlnGlyIleAla 944	
QY	787	CTCCAGCAAGCACCTCTAGGTATACCAGCA-----CCCTGTTTC 825	
DB	945	GlyGlnArgGlyValValGlyLeuProGlyGlnArgGlyGluArgGlyPheProGlyLeu 964	
QY	826	CCTTCGGCGGGGAAGTCATAGTTAGGGAGCACCTCACACTTTCCTCAGGAGAACACACA 885	
DB	965	ProGlyProSerGlyGluProGlyLysGlnGlyProSerGlyAlaSerGlyGluArgGly 984	
QY	886	CAACCATGACCTCAGCTGGGAGACCCCAACTCCAGCCCTCCAGCCCAACCTTGCCCA 945	

DB	985	ProProGlyPro-----MetGlyProProGlyLeuAlaGlyProProGly-----Glu 1000	
QY	946	GCCAGCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCTAGGAACCCACGGGAAGTAC 1005	
DB	1001	AlaGlyArgGluGlyAla-----ProGlyAlaGlu 1010	
QY	1006	TGTGCTCCAGGAGAGAGCTGGAGTTCTTGACACACAGCAAGCGTGGTGGCTGGTAA 1065	
DB	1011	GlyAlaProGlyArgAsp---GlyAlaAlaGlyProLys----- 1022	
QY	1066	GAATGAGGGGACGAGCGGCTACATCTCCCAAGCAACATCTCTGGAGCCCTACACCGGG 1125	
DB	1023	-----GlyAspArgGlyGluThrGlyProAlaGlyProProGlyAlaPro-----Gly 1038	
QY	1126	GACCCCTGGGACCCAGGCGCTACCTCTCGGGTTCAATGCTTCGACTTAGCTCGAG 1185	
DB	1039	AlaProGlyAlaProGlyProValGlyProAlaGlyLysAsnGlyAsp----- 1054	
QY	1186	GCCTGAAGAGGTACAGACTGGCTGCAGGCGAGAGAACTTCTCCACTGCCACGGTAGGAC 1245	
DB	1055	-----ArgGlyGluThrGlyProAlaGly----- 1062	
QY	1246	ACTTGGTCCCTAGCGGGAGGCCACTTCTCGCATAGACCTGGGAGCTACAGTCT 1305	
DB	1063	-----ProAlaGlyProPro-----GlyProAlaGlyAla 1072	
QY	1306	ATGTCACAGAGGAGCCCAAGCAATCTGTCCGGCTGGAGGCTGTCAAGAGTGTCTGG 1365	
DB	1073	ArgGlyProAlaGlyProGlnGlyProArgGlyAspLysGlyLeuThrGlyGluGlnGly 1092	
QY	1366	GATAAGCCTTAGGCACCACTTACAGCTTCCAGAACAGCCCGCTGTATGCAAGAT 1425	
DB	1093	AspArgGlyMetLysGlyHisArgGlyPheSerGlyLeuGlnGlyProProGlyProPro 1112	
QY	1426	GCCACATCTGATACCCATAGAGCCCC---GAGATTCCTCTTCTGGATCCAGTTGCA 1482	
DB	1113	Gly-----AlaProGlyGluGlnGlyProSerGlyAlaSerGlyPro 1126	
QY	1483	GCAAACCCACACCCAGCTCACAGCAAAACAATGCAGACAGGCGCAGAGCTGAAGCA 1542	
DB	1127	AlaGlyPro-----ArgGlyProProGlySerAla 1136	
QY	1543	-----AACAGTGTCTTCTGGCTGTGGTGGAGCCCTCCC 1576	
DB	1137	GlyAlaAlaGlyLysAspGlyLeuAsnGlyLeuPro---GlyProIleGly-ProPro 1154	
RESULT 15			
ID	CALL_MOUSE	STANDARD:	PRT: 1453 AA.
AC	P11087; Q60635;		
DT	01-JUL-1989 (Rel. 11, Created)		
DT	01-NOV-1997 (Rel. 35, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Collagen alpha 1(I) chain precursor.		
GN	COL1A1 OR COL1A1.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN	SEQUENCE FROM N.A.		
RP	STRAIN=FVB/N;		
RC	MEDLINE=96033240; PubMed=8535610;		
RA	Li S.W., Khillan J., Prockop D.J.;		
RT	"The complete cDNA coding sequence for the mouse pro alpha 1(I) chain		
RT	of type I procollagen.";		
RL	Matrix Biol. 14:593-595(1995).		
RN	[2]		
RP	SEQUENCE OF 518-1128 FROM N.A.		
RC	MEDLINE=86137403; PubMed=3841523;		
RA	French B.T., Lee W.-H., Maul G.G.;		
RT	"Nucleotide sequence of a cDNA clone for mouse pro alpha 1(I)		
RT	collagen protein.";		

Gene 39:311-312(1985).
[3]
SEQUENCE OF 735-1130 FROM N.A.
MEDLINE=83141374; PubMed=6298597;
RA Monson J.M., Friedman J., McCarthy B.J.;
RT "DNA sequence analysis of a mouse pro alpha 1(I) procollagen gene:
evidence for a mouse B1 element within the gene.";
RL Mol. Cell. Biol. 2:1362-1371(1982).
[4]
SEQUENCE OF 735-878 AND 1005-1058 FROM N.A.
MEDLINE=83157109; PubMed=6219867;
RA Monson J.M., McCarthy B.J.;
RT "Identification of a half/c mouse pro alpha 1(I) procollagen gene:
evidence for insertions or deletions in gene coding sequences.";
RL DNA 1:59-69(1981).
[5]
SEQUENCE OF 1442-1453 FROM N.A.
MEDLINE=88124276; PubMed=3340560;
RA Moolenaar K., Harbers K.;
RT "Two mRNAs of mouse pro alpha 1(I) collagen gene differ in the size
of the 3'-untranslated region.";
RL Nucleic Acids Res. 16:773-773(1988).
CC -1- FUNCTION: TYPE I COLLAGEN IS A MEMBER OF GROUP I COLLAGEN
(FIBRILLAR FORMING COLLAGEN).
CC -1- SUBUNIT: TRIMERS OF ONE ALPHA 2(I) AND TWO ALPHA 1(I) CHAINS.
CC -1- TISSUE SPECIFICITY: FORMS THE FIBRILS OF TENDON, LIGAMENTS AND
BONES. IN BONES THE FIBRILS ARE MINERALIZED WITH CALCIUM
HYDROXYAPATITE.
CC -1- PTM: PROLINES AT THE THIRD POSITION OF THE TRIPEPTIDE REPEATING
UNIT (G-X-Y) ARE HYDROXYLATED IN SOME OR ALL OF THE CHAINS.
CC -1- SIMILARITY: CONTAINS 1 VWFC DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC
DR EMBL; U08020; AAA88912.1; -
DR EMBL; X15896; CAA33904.1; -
DR EMBL; M14423; AAA37333.1; -
DR EMBL; M17491; AAA37334.1; -
DR EMBL; K06753; CAA29927.1; -
DR EMBL; K03036; AAA37332.1; -
DR EMBL; K03029; AAA37332.1; JOINED.
DR EMBL; K03030; AAA37332.1; JOINED.
DR EMBL; K03031; AAA37332.1; JOINED.
DR EMBL; K03032; AAA37332.1; JOINED.
DR EMBL; K03033; AAA37332.1; JOINED.
DR EMBL; K03034; AAA37332.1; JOINED.
DR EMBL; K03035; AAA37332.1; JOINED.
DR PIR; A23982; A23982.
DR MGD; MGI:88467; Colla1.
DR InterPro; IPR000087; Collagen.
DR InterPro; IPR000885; Fib.collagen_C.
DR InterPro; IPR001007; VWFC_C.
DR Pfam; PF01391; Collagen; 18.
DR Pfam; PF01410; COLFI; 1.
DR ProDom; PD000007; Collagen; 1.
DR ProDom; PD002078; Fib.collagen_C; 1.
DR SMART; SM00038; COLFI; 1.
DR SMART; SM00214; VWFC; 1.
DR PROSITE; PS01208; VWFC; 1.
KW Extracellular matrix; Connective tissue; Repeat; Hydroxylation;
KW Glycoprotein; Collagen; Signal.
FT SIGNAL 1 22
FT PROPEP 23 151 AMINO-TERMINAL PROPEPTIDE.
FT CHAIN 152 1207 COLLAGEN ALPHA 1(I) CHAIN.
FT PROPEP 1208 1453 CARBOXYL-TERMINAL PROPEPTIDE.
FT DOMAIN 29 87 VWFC.
FT DOMAIN 152 167 NONHELICAL REGION (N-TERMINAL).
FT

FT	DOMAIN	168	1181	TRIPLE-HELICAL REGION.
FT	DOMAIN	1182	1207	NONHELICAL REGION (C-TERMINAL).
FT	CARBOHYD	56	56	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	1354	1354	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	SITE	734	736	CELL ATTACHMENT SITE (POTENTIAL).
FT	SITE	1082	1084	CELL ATTACHMENT SITE (POTENTIAL).
FT	CONFLICT	1450	1450	A -> V (IN REF. 5).
SQ	SEQUENCE	1453 AA;	137944 MW;	3B802E535DF81808 CRC64;

Alignment Scores:
Pred. No.: 0.00227 Length: 1453
Score: 180.00 Matches: 147
Percent Similarity: 32.93% Conservative: 42
Best Local Similarity: 25.61% Mismatches: 226
Query Match: 5.76% Indels: 159
DB: 1 Gaps: 35

US-09-762-021A-1 (1-1710) x CALL_MOUSE (1-1453)

Qy	1	GGCAGAGCGACTGAAGACCGCCTGCA	-----GAAGGCTCTGGAGGAAGA	45
Db	618	GlyGluArgGlyGluGlnGlyProAlaGlySerProGlyPheGlnGlyLeuProGlyPro	637	
Qy	46	GCTGGAGCAAGACCTCGACTTGGAGCGCTTACGACGAG	-----CCAGGACAGATGGAG	99
Db	638	AlaGlyProGlyGluAlaGlyLysProGlyGluGlnGlyValProGlyAspLeu	656	
Qy	100	GGGCGCTGCTATGGAAGCGCGCTCCCTATGAGCA	-----	135
Db	657	-----GlyAlaProGlyProSerGlyAlaArgGlyGluArgGlyPheProGly	672	
Qy	136	-----GGCAGCGCTATCTGGAGCGCGGATCCCTCCAGAACACGCCACCA	180	
Db	673	GluArgGlyValGlnGlyProGlyProGlyProAlaGlyProArgGlyAsnAsnGlyAlaPro	692	
Qy	181	GAGGACCTAGAGCAGCAGCTCCACC	-----ATCCCCAAGGCCCTGCCAGCCACACAG	237
Db	693	GlyAsnAspGlyAlaLysGlyAspThrGlyAlaProGlyAlaProGlySer	711	
Qy	238	TGCCCGAGAACCAAGTCCCTTACTCTGCTCCCTCCCAAGGCGGCTCTTCCCTCCCGAGGA	297	
Db	712	AlaProGlyLeuGlnGlyMetProGlyGluArgGlyAlaAlaGlyLeu	729	
Qy	298	CCAGAGGAGGAGGAGGAGGAGTGTGCTGAACATGCTCTAAGGACACATTGAGCTGTTTCATGGG	357	
Db	730	ProLysGlyAspArgGlyAspAlaGly	-----ProLysGly	741
Qy	358	AAAGCTGGAGAGGCCCGCAGGCAAGACAGCAGCAGGAGAGAAATTTGGGAAAAAAACAA	417	
Db	742	---AlaAspGlySerProGlyLysAspGlyAlaArg	-----	752
Qy	418	GGACCGGAGGAGTCTACCCAGGACAGTACATTGCTTCCAGAGATCAAGTACAG	477	
Db	753	-----GlyLeuThrGlyPro	-----	757
Qy	478	CTTCAACCTCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAGTCCCTCGAGCTCGT	537	
Db	758	IleGlyProGlyProAlaGly	-----AlaProGlyAspLysGlyGluAlaGlyPro	775
Qy	538	ACATACCTCTTCAAGTCCCTGAACCTTATCTCTGGCCAGGTCGCCCTGAGGCTGGCCCTAGC	597	
Db	776	SerGlyPro	-----ProGly	786
Qy	598	AGCCCAAGTGTAT	-----CTCACCCCTCTCCACCCCTAAAGCTAT	636
Db	787	AlaProGlyAspArgGlyGluAlaGlyProGlyProAlaGlyPheAlaGlyProPro	806	
Qy	637	-----CAACCTGTACAGTCTGTCTAAGCCACCTCCAGAGTAACCTTTTGGAT	684	
Db	807	GlyAlaAspGlyGlnProGly	-----AlaLysGlyGlu	825
Qy	685	GGGTTTGGCCCGAGCTGGACCACTAGCGGGCGGCTGGAGCGGCGATGAGCCCTGCC	744	

Db 825 yAspAlaGlyPro-ProGlyPro---AlaGlyProAlaGly-----p 838
QY 745 CTACCAACCCACATCTCAGATGAGTGGCAACTTCCAGAGCCCTCCAGCCAAAGCACCTT 804
Db 838 roProGlyPro-----IleGlyAsnValGlyAlaProGlyProLysGlyProA 854
QY 805 AGGATACC-----AGGACCTCTTTCCCTTCGGCGGGGAAAGTCATAGGTT 849
Db 854 rgGlyAlaAlaGlyProProGlyAlaThrGlyPheProGlyAlaAlaGlyArgValGlyP 874
QY 850 AGGGAGCACCTCACACTTTCCTCAGGAGAGACACACAACCATGACCTCAGCCTGGGGA 909
Db 874 roProGlyPro-----S 878
QY 910 CCCCAACTCCAGGCCCTCCAGCCGCCCAAACTGCCAGCCAGCCCTGAAATGCAAGTCTT 969
Db 878 erGlyAsnAlaGlyProProGlyPro-ProGlyProValGlyLysGlyLysLys--- 896
QY 970 GTACGAGTTTGAAGCTAGGAACCCACGGGAACTGACTGTGTGTC---GGGAGA----- 1020
Db 897 -----GlyProArgGly---GluThrGlyProAlaGlyArgProGly 909
QY 1021 GAAGCTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGTGAAGAATGAGCGGGGAGG 1080
Db 910 GluValGlyProProGlyProProGlyPro-----AlaGlyGlu----- 922
QY 1081 GAGCGGCTACATTCCCAAGCAACATCTCGAGCCCT---ACAGCCGGGACCCCTGGGAC 1137
Db 923 -----LysGlySerProGlyAlaAspGlyProAlaGlySerProGlyThr 937
QY 1138 CCAGGCCAGTCAACCTTCGGGTTCCAATGCTTCGACTTACGAGGCCCTGAAGAGGT 1197
Db 938 ProGlyProGlnGlyIleAlaGlyGlnArgGlyValValGlyLeuProGlyGlnArgGly 957
QY 1198 CACAGACTGGCTGCAGGCGAGAACTTCTCCACTGCCCGGTGAGGACACTTGGGTCCCT 1257
Db 958 GluArgGlyPheProGlyLeuPro-----GlyPro 967
QY 1258 GACGGGGAGCCAGCTACTTCGCATAAGACCTGGGAGCTACAGATGCTATGTCCACAGGA 1317
Db 968 SerGlyGluProGlyLysGlnGlyProSer---GlySerSerGlyGluArgGlyProPro 986
QY 1318 GGCCCCACGAATCTCTCCCGCTGGAGGCTGTCAGAAGGATGCTGGG---GATAAGCCC 1374
Db 987 GlyProMetGlyProProGlyLeuAlaGlyProProGlyGluSerGlyArgGluGlySer 1006
QY 1375 TTAGGCACCACTTAGACACCTCCCAAGAACACAGGCCCGCTGATGCAAGATGGCAGATCT 1434
Db 1007 ProGlyAlaGluGlySerProGlyArgAspGlyAlaProGlyAlaLysGlyAspArgGly 1026
QY 1435 GATACCCATTAGAGCCCGCCGAGAATCTCTCTGATCCAGTTTGCAGCAAAACCCACA 1494
Db 1027 GluThrGlyProAlaGlyProProGlyAlaProGlyAlaProGlyAlaProGlyProVal 1046
QY 1495 CCCCAGCTCACACAG-----CAAAACAATGGACAGAGGCCGAGCAAGCAACAGT 1548
Db 1047 GlyProAlaGlyLysAsnGlyAspArgGlyGluThrGlyProAlaGlyProAlaGlyPro 1066
QY 1549 GTC---CCTTCGGCTGTGTGA---GCCTCCCCA 1578
Db 1067 IleGlyProAlaGlyAlaArgGlyProAlaGlyPro 1078

Search completed: February 25, 2003, 19:41:10
Job time : 53.5 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 17:00:28 ; Search time 53 Seconds

(without alignments)

13295.870 Million cell updates/sec

Title: US-09-762-021a-1

Perfect score: 3127

Sequence: 1 ggcagagcgactgaagacca.....aaaaaaaaaataangataaa 1710

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Zgapop 6.0 , Zgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO.spool/US09762021/runat_24022003_153038_7557/app.query.fasta_1.1863
-DB=SPTREMBL_21 -OFMT=fastan -SUFFIX=rspt -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM-ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09762021 -ECGN_1_1_63 @runat_24022003_153038_7557 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAPP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL_21.*
1: sp_archaea.*
2: sp_bacteria.*
3: sp_fungi.*
4: sp_human.*
5: sp_invertebrate.*
6: sp_mammal.*
7: sp_mhc.*
8: sp_organelle.*
9: sp_phage.*
10: sp_plant.*
11: sp_rodent.*
12: sp_virus.*
13: sp_vertebrate.*
14: sp_unclassified.*
15: sp_rviro.*
16: sp_bacteriap.*
17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	ID	Description
1	2416.5	77.3	594	4	Q96E47	Q96e47 homo sapien

2	2416	77.3	593	4	Q8TE67	Q8te67 homo sapien
3	2229	71.3	563	4	Q9H719	Q9h719 homo sapien
4	1606.5	51.4	600	11	Q91WL0	Q91wl0 mus musculu
5	597.5	19.1	729	11	Q9K30	Q9k30 mus musculu
6	588	18.8	743	4	Q9H6S3	Q9h6s3 homo sapien
7	584	18.7	715	4	Q9H8K9	Q9h8k9 homo sapien
8	574.5	18.4	596	4	Q96BB7	Q96bb7 homo sapien
9	572.5	18.3	723	4	Q8TE68	Q8te68 homo sapien
10	571.5	18.3	596	4	Q9NXH0	Q9nxh0 homo sapien
11	563	18.0	652	11	Q9D2M6	Q9d2m6 mus musculu
12	563	18.0	716	11	Q8R5F8	Q8r5f8 mus musculu
13	403	12.9	778	5	Q9VP07	Q9vp07 drosophila
14	402	12.9	778	5	Q9U3Z9	Q9u3z9 drosophila
15	399	12.8	437	4	Q9GZQ2	Q9gsq2 homo sapien
16	386	12.3	505	4	Q9BSQ2	Q9bsq2 homo sapien
17	356	11.4	929	5	O18249	O18249 caenorhabdi
18	352	11.3	732	5	O18250	O18250 caenorhabdi
19	351	11.2	712	5	Q9STJ6	Q9stj6 drosophila
20	341.5	10.9	714	5	Q9VEP7	Q9vep7 drosophila
21	333.5	10.7	472	11	Q8R0D6	Q8r0d6 mus musculu
22	329.5	10.5	327	4	Q9BUN9	Q9bum9 homo sapien
23	327.5	10.5	327	4	Q96K06	Q96k06 homo sapien
24	276.5	8.8	232	4	Q8WYW7	Q8wyw7 homo sapien
25	215	6.9	400	11	Q91VT7	Q91vt7 mus musculu
26	200.5	6.4	3084	12	Q8U211	Q8u211 pseudorabie
27	196.5	6.3	890	12	Q8UZE1	Q8uze1 cercopithic
28	196.5	6.3	1548	4	Q9NYW9	Q9nyw9 homo sapien
29	196.5	6.3	2161	4	Q9Y566	Q9y566 homo sapien
30	196.5	6.3	5147	4	Q9Y6V0	Q9y6v0 homo sapien
31	195.5	6.3	3926	4	Q9UPA5	Q9upa5 homo sapien
32	195	6.2	645	4	Q96EX0	Q96ex0 homo sapien
33	195	6.2	1985	5	Q8T9N4	Q8t9n4 drosophila
34	193	6.2	606	11	Q8R0Y2	Q8r0y2 mus musculu
35	192.5	6.2	847	10	Q9XIB6	Q9xib6 arabidopsis
36	192	6.1	2167	11	Q9WV48	Q9wv48 rattus norv
37	191.5	6.1	437	12	Q69146	Q69146 human herpe
38	191	6.1	1300	12	O36421	O36421 alcelaphine
39	191	6.1	1963	5	Q9VSK5	Q9vsk5 drosophila
40	190.5	6.1	925	5	Q93238	Q93238 caenorhabdi
41	189.5	6.1	2158	11	Q9WU13	Q9wu13 rattus norv
42	189	6.0	1188	10	Q41805	Q41805 zea mays (m
43	189	6.0	2087	11	Q9W0E8	Q9wue8 rattus norv
44	188	6.0	1445	13	Q93251	Q93251 rana catesb
45	187	6.0	608	12	Q9Q5K9	Q9q5k9 herpesvirus

ALIGNMENTS

RESULT 1

Q96E47	Q96E47	PRELIMINARY;	PRT;	594 AA.
ID	Q96E47	PRELIMINARY;	PRT;	594 AA.
AC	Q96E47			
DT	01-DEC-2001	(TrEMBLrel. 19, Created)		
DT	01-DEC-2001	(TrEMBLrel. 19, Last sequence update)		
DT	01-JUN-2002	(TrEMBLrel. 21, Last annotation update)		
DE	Similar to hypothetical protein FLJ21522.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=COLON;			
RA	Strausberg R;			
RL	Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; BC012926; AAH12926.1; -			
DR	InterPro; IPR000050; PID_domain.			
DR	InterPro; IPR001452; SH3.			
DR	Pfam; PF00018; SH3; 1.			
DR	ProDom; PD000066; SH3; 1.			
DR	SMART; SM00462; PTB; 1.			
DR	PROSITE; PS00002; SH3; 1.			
SQ	SEQUENCE 594 AA; 66974 MW; F71E8F9B7564DEFE CRC64;			

Alignment Scores:

Pred. No.: 3,88e-184 Length: 594
Score: 2416.50 Matches: 458
Percent Similarity: 99.78% Conservative: 0
Best Local Similarity: 99.78% Mismatches: 0
Query Match: 77.28% Indels: 1
DB: 4 Gaps: 1

US-09-762-021A-1 (1-1710) x Q96E47 (1-594)

```
QY 2 GCAGAGCGACTGAAGACCAAGCTGAGAGGCTCTGGAGAGAGAGCTGGAGCAA---AGA 58
DB 136 AlaGluArgLeuLysThrSerLeuGlnLysAlaLeuGluGluGluGlnSerArg 155
QY 59 CCGTGAAGTGGAGGCTTCAGCCAGCCAGGACAGATGAGGGGCGCTTATGGAAAGG 118
DB 156 ProArgLeuGlyGlyLeuGlnProGlyGlnAspArgTrpArgGlyProAlaMetGluArg 175
QY 119 CCGCTCCCTATGGAGCAGCAGCTATCTGGAGCGCGGGATCCCTCCAGAACAGCCAC 178
DB 176 ProLeuProMetGluGlnAlaArgTyrLeuGluProGlyLeProGluGlnProHis 195
QY 179 CAGAGACCCCTAGACACAGCTCCACCATCCCAAGGCCCTGCCAGCCACACCACT 238
DB 196 GlnArgThrLeuGluHisSerLeuProProSerProArgProArgHisThrSer 215
QY 239 GCCCGAGAACAAGTCCCTTTACTCTGCTCTCTCCAAAGCGCTCTTCTCCCGGAGGAC 298
DB 216 AlaArgGluProSerAlaPheThrLeuProProProArgSerSerProGluAsp 235
QY 299 CCAGAGAGGAGCAGGAGTGTCTGAACCATGTCTTAAGGACATTCAGCTGTTCATGGGA 358
DB 236 ProGluArgAspGluGluValLeuAsnHisValLeuArgAspGluLeuPheMetGly 255
QY 359 AGCTGGAGAAGCCCGAGCAAGACCCAGCAGAGAGAAATTTGGGAAAAAACACAG 418
DB 256 LysLeuGluLysAlaGlnAlaLysThrSerArgLysLysLysPheGlyLysLysAsnLys 275
QY 419 GACCAGGAGGTCTCACCCAGCAGACATGACTGCTCCAGAGATCAAGTACAGC 478
DB 276 AspGlnGlyLeuThrGlnAlaGlnTyrIleAspCysPheGlnLysIleLysTyrSer 295
QY 479 TTCAACCTCTCTGGAGGCTGCCACCTGGCTGAAGAGACAAAGTCCCTCAGCTCGTA 538
DB 296 PheAsnLeuLeuGlyArgLeuAlaThrTrpLeuLysGluThrSerAlaProGluLeuVal 315
QY 539 CACATCCTCTCAAGTCCCTGAATTCATCTCTGGCCAGGTGCCCTGAGGCTGGCTAGCA 598
DB 316 HisIleLeuPheLysSerLeuAsnPheIleLeuAlaArgCysProGluAlaGlyLeuAla 335
QY 599 GCCCAAGTATCTCACCCCTCTCACCCCTAAAGCTATCAAGCTGCTACAGTCTGTCTA 658
DB 336 AlaGlnValIleSerProLeuLeuThrProLysAlaIleAsnLeuLeuGlnSerCysLeu 355
QY 659 AGCCCACTCTGAGAGTAACCTTTGGATGGGCTGGGCCAGCCCTGAGCAGCCGGGC 718
DB 356 SerProGluSerAsnLeuTrpMetGlyLeuGlyProAlaTrpThrSerArgAla 375
QY 719 GACTGGACAGGGATGAGGCCCTGCCCTACCAACCCACATTTCTCAGATGACGTGCAACT 778
DB 376 AspTrpThrGlyAspGluProLeuProTyrGlnProThrPheSerAspAspTrpGlnLeu 395
QY 779 CCAGAGCCCTCCAGCAACCCCTTAGCATACAGGACCCCTGTTCCCTTCGGCGGGA 838
DB 396 ProGluProSerSerGlnAlaProLeuGlyTyrGlnAspProValSerLeuArgGly 415
QY 839 ACTCATAGCTTGGAGCAGCTCACACTTTCTCTCAGGAGAGACACACACCAACCATGACCT 898
DB 416 SerHisArgLeuGlySerThrSerHisPheProGlnGluLysThrHisAsnHisAspPro 435
QY 899 CAGCTGGGAGCCCAACTCCAGGCGCTCCAGCCCAACCTGCGCCAGCCACCTCGAAA 958
```

```
DB 436 GlnProGlyAspProAsnSerArgProSerSerProLysProAlaGlnProAlaLeuLys 455
QY 959 ATGCAAGTCTTACGAGTTTCAAGCTAGGAACCCACGGGAACCTGACTGCTGCTCCAGGA 1018
DB 456 MetGlnValLeuTyrGluPheGluAlaArgAsnProArgGluLeuThrValGlnGly 475
QY 1019 GAGAAGCTGGAGTCTCTGACACACAGCGGTGGTGGTGAAGAAATGAGCGGGA 1078
DB 476 GluLysLeuGluValLeuAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGly 495
QY 1079 CGGACGGCTACATTCACAGCAACATCTGGAGCCCTACACCCGGGGACCCCTGGGACC 1138
DB 496 ArgSerGlyTyrIleProSerAsnIleLeuGluProLeuGlnProGlyThrProGlyThr 515
QY 1139 CAGGCGCAGTCAACCTCTCGGTTTCAATGCTTCCACTTAGCTGAGGCTCAAGAGGTC 1198
DB 516 GlnGlyGlnSerProSerArgValProMetLeuArgLeuSerSerArgProGluGluVal 535
QY 1199 ACAGACTGGCTCAGCAGCAGAACTTCTCCACTGCCAGCGTGAGGACACTTGGGTCCCTG 1258
DB 536 ThrAspTrpLeuGlnAlaGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeu 555
QY 1259 ACGGGGAGCCAGCTACTTCGCATAGACCTGGGAGCTACAGATGCTATGTCACAGGAG 1318
DB 556 ThrGlySerGlnLeuArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGlu 575
QY 1319 GCCCCAGCAATCTCTCCCGCTGGAGGCTGTCAAGAGATGCTGGGATAAGCCCT 1375
DB 576 AlaProArgIleLeuSerArgLeuGluAlaValArgArgMetLeuGlyIleSerPro 594
```

RESULT 2

```
Q8TE67 PRELIMINARY: PRT: 593 AA.
AC Q8TE67:
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor pathway substrate 8 related protein
DE 3.
GN EPSR83
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Scita G., Confalonieri S., Offenhauser N., Borgonovo A., Tocchetti A.,
RA Romano P., Di Fiore P.P.;
RT "Cloning and characterization of novel members of the Eps8 protein
RT family.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY074930; AAL76119.1; -
KW Receptor.
SQ SEQUENCE 593 AA; 66891 MW; E33134671B9150AF CRC64;
```

Alignment Scores:
Pred. No.: 4,25e-184 Length: 593
Score: 2416.00 Matches: 456
Percent Similarity: 99.78% Conservative: 1
Best Local Similarity: 99.56% Mismatches: 1
Query Match: 77.26% Indels: 0
DB: 4 Gaps: 0

US-09-762-021A-1 (1-1710) x Q8TE67 (1-593)

```
QY 2 GCAGAGCGACTGAAGACCAAGCTGAGAGGCTCTGGAGAGAGAGCTGGAGCAAGACCT 61
DB 136 AlaGluArgLeuLysThrSerLeuGlnLysAlaLeuGluGluGluLeuGlnArgPro 155
QY 62 CGACTTGGAGGCTTCAGCCAGGACAGATGAGGGGCGCTGCTATGCAAGAGCGG 121
DB 156 ArgLeuGlyGlyLeuGlnProSerGlnAspArgTrpArgGlyProAlaMetGluArgPro 175
```



```
QY 122 CTCCTATGAGCAGGACGCTATCTGGAGCGGGATCCCTCCAGAACAGCCCGACAG 181
Db 176 LeuProMetGluGlnAlaAargTyrLeuGluProGlyIleProProGluGlnProHisGln 195
QY 182 AGGACCTTAGACACAGCCTCCACCATCCCAAGCGCCCTGCCACAGCCACACAGTGC 241
Db 196 ArgThrLeuGluHisSerLeuProProSerProArgProLeuProArgHisThrSerAla 215
QY 242 CCAGAACCAAGTCTTACTCTGCTCTCTCCAAAGCGGTCTCTCTCCCGGAGGACCCA 301
Db 216 ArgGluProSerAlaPheThrLeuProProProArgArgSerSerProGluAspPro 235
QY 302 GAGAGGCGAGGAAGTCTGAACCATGCTCTTAAGGAGCATTGAGCTGTTCATGGGAAG 361
Db 236 GluArgAspGluGluValLeuAsnHisValLeuArgAspIleGluLeuPheMetGlyLys 255
QY 362 CTGGAGAGGCCCGCGCAAGACAGCAGGAGAGAAATTTGGGAAAAAACAAGGAC 421
Db 256 LeuGluLysAlaGlnAlaLysThrSerArgLysLysLysPheGlyLysAsnLysAsp 275
QY 422 CAGGAGGTCTCACCGCAGCAGTACATTGACTCTTCCAGAAGCATCAAGTACAGCTTC 481
Db 276 GlnGlyGlyLeuThrGlnAlaGlnTyrIleAspCysPheGlnLysIleLysHisSerPhe 295
QY 482 AACCTCTGGGAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCTGAGCTCGTACAC 541
Db 296 AsnLeuLeuGlyArgLeuAlaThrTrpLeuLysGluThrSerAlaProGluLeuValHis 315
QY 542 ATCTCTTCAAGTCTCAACTTCATCTGCGCCAGGTGCCCTGAGCGTGGCTAGCAGCC 601
Db 316 IleLeuPheLysSerLeuAsnPheIleLeuAlaArgCysProGluAlaGlyLeuAlaAla 335
QY 602 CAAGTGATCTCACCCCTCTCAACCCTAAAGCTATCAACCTGCTACAGTCTCTTAAGC 661
Db 336 GlnValIleSerProLeuLeuThrProLysAlaIleAsnLeuLeuGlnSerCysLeuSer 355
QY 662 CCACCTGAGAGTAACCTTTGGATGGGTGGGCCCGAGCTGGACACTAGCGGGCCGAC 721
Db 356 ProProGluSerAsnLeuTrpMetGlyLeuGlyProAlaTrpThrThrSerArgAlaAsp 375
QY 722 TGGACAGCGGATGACCCCTGCCCTACCAACCCCATCTCTAGATGACTGGCACTTCCA 781
Db 376 TrpThrGlyAspGluProLeuProTyrGlnProThrPheSerAspTrpGlnLeuPro 395
QY 782 GAGCCCTCGAGCAAGCACCTTAGGATACAGGACCTGTTTCCCTTCGGCGGGAAGT 841
Db 396 GluProSerSerGlnAlaProLeuGlyTyrGlnAspProValSerLeuArgArgGlySer 415
QY 842 CATAGGTTAGGAGACCTCACACTTCTCTCAGGAGAGACACACACCATGACCCCTCAG 901
Db 416 HisArgLeuGlySerThrSerHisPheProGlnGluLysThrHisAsnHisAspProGln 435
QY 902 CTTGGGGACCCCACTCCAGGCCCTCCAGCCCAACCTGCCACGACCCCTGAAATG 961
Db 436 ProGlyAspProAsnSerArgProSerSerProLysProAlaGlnProAlaLeuLysMet 455
QY 962 CAAGTCTTGTACAGTTTGAAGCTAGGAACCCACGGGAACCTGACTGTGTCTCAGGAGAG 1021
Db 456 GlnValLeuTyrGluPheGluAlaArgAsnProArgGluLeuThrValValGlnGlyGlu 475
QY 1022 AAGCTGGAGGTTCCTGGACACAGCAAGCGGTGGTGGCTGGTGAAGAATCAGCGGGACGG 1081
Db 476 LysLeuGluValLeuAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArg 495
QY 1082 ACGGCTACATCCCAAGCAACATCTCGAGGCCCTTACAGCGGGGACCCCTGGGACCCAG 1141
Db 496 SerGlyTyrIleProSerAsnIleLeuGluProLeuGlnProGlyThrProGlyThrGln 515
QY 1142 GCCAGTCAACCTCTCTGGGTCCAATGCTTCGACTTAGCTCGAGGCCCTGAAGAGTGCACA 1201
Db 516 GlyGlnSerProSerArgValPrometLeuArgLeuSerSerArgProGluValThr 535
QY 1202 GACTGGCTGCAGGCGAGAACTTCTCCACTGCCACGGTGAGGACACTTGGGTCTCGACG 1261
```

```
Db 536 AspTpLeuGlnAlaGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThr 555
QY 1262 GGGAGCCAGCTACTTCGCATAGACCTGGGAGCTTACAGATGCTATGTCCACAGAGGCC 1321
Db 556 GlySerGlnLeuLeuArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAla 575
QY 1322 CCACCAATCTGTCTCCGCTGGAGGCTGTCTCAGAAGGATGCTGGGATAAGCCCT 1375
Db 576 ProArgIleLeuSerArgLeuGluAlaValArgArgMetLeuGlyIleSerPro 593
RESULT 3
Q9H719 PRELIMINARY; PRT; 563 AA.
AC Q9H719;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)
DE CDNA: FLJ21522 fis, clone COL05884.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Kавabата А., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Ohayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AK0251175; BAB15081.1; -.
DR HSSP; Q08509; 1AQJ
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00462; PTB; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 563 AA; 63527 MW; A36915ED19057E9B CRC64;
Alignment Scores:
Pred. No.: 3,36e-169 Length: 563
Score: 2229.00 Matches: 428
Percent Similarity: 93.45% Conservative: 0
Best Local Similarity: 93.45% Mismatches: 0
Query Match: 71.28% Indels: 30
Db: 4 Gaps: 1
US-09-762-021A-1 (1-1710) x Q9H719 (1-563)
QY 2 GCAGAGCGACTCAAGCAGCAGCTCGCAGAGGCTCGAGAGAGCTGGAGAAAGCTGGAGCAAGACCT 61
Db 136 AlaGluArgLeuLysThrSerLeuGlnLysAlaLeuGluGluLeuGluGlnArgPro 155
QY 62 CGACTTGGAGGCTTTCAGCCAGCCAGGACAGATGGAGGGGCGCTGCTATGTAAGAGCGCG 121
Db 156 ArgLeuGlyLeuGlnProGlyGlnAspArgTrpArgGlyProAlaMetGluArgPro 175
QY 122 CTCCTATGAGCAGCAGCTATCTGGAGCGGGATCCCTCCAGAACAGCCCGACAG 181
Db 176 LeuProMetGluGlnAlaArgTyrLeuGluProGlyIleProProGluGlnProHisGln 195
QY 182 AGGACCTTAGACACAGCCTCCACCATCCCAAGCGGTCTCTCTCCCGGAGGACCCA 241
Db 196 ArgThrLeuGluHisSerLeuProProSerProArgProLeuProArgHisThrSerAla 215
QY 242 CCAGAACCAAGTCTTACTCTGCTCTCTCCAAAGCGGTCTCTCTCCCGGAGGACCCA 301
```

```
Db 216 ArgGluProSerAlaPheThrLeuProProArgArgSerSerSerProGluAspPro 235
QY 302 GAGAGGACGAGAGTGTGTAACCATGCTCTAAAGGACATTGAGCTGTTCATGGAAAG 361
Db 236 GluArgAspGluGluValLeuAsnHisValLeuArgAspIleGluLeuPheMetGlyLys 255
QY 362 CTGGAGAAGCCCGAGCAAGACACGACGAGAGAAATTTGGGAAAAAACAAGGAC 421
Db 256 LeuGluLysAlaGlnAlaLysThrSerArgLysLysLysPheGlyLysAsnLysAsp 275
QY 422 CAGGGAGGTCTCACCAGGACGACAGTACATTGACTGCTCCAGAAGATCAAGTACACTTC 481
Db 276 GlnGlyLeuThrGlnAlaGlnIleAspCysPheGlnLysIleLysIleLysIleLysIle 295
QY 482 AACCTCTGGGAGGTGGCCACCTGGCTGAAGGACAAAGTGCCTCAGCTCGGTACAC 541
Db 296 AsnLeuLeuGlyArgLeuAlaThrTrpLeuLysGluThrSerAlaProGluLeuValHis 315
QY 542 ATCTCTTCAAGTCCCTGAACCTTCATCTGCGCCAGGTGCCCTGAGGCTGCCCTAGCACC 601
Db 316 IleLeuPheLysSerLeuAsnPheIleLeuAlaArgCysProGluAlaGlyLeuAlaAla 335
QY 602 CAAGTATCTACCCCTCTCACCCTTAAGCTATCAAGCTATCAAGCTGCTACAGTCTGTAAAC 661
Db 336 GlnValIleSerProLeuLeuThrProLysAlaIleAsnLeuLeuGlnSerCysLeuSer 355
QY 662 CCACCTGAGAGTAACCTTTGGATGGGTGGGCCAGCTGGACCACTAGCCGGCCGAC 721
Db 356 ProProGluSerAsnLeuTrpMetGlyLeuGlyProAlaTrpThrSerArgAlaAsp 375
QY 722 TGGACAGGCGATGAGCCCTGCCCTACCAACCCACATCTCAGATGACTGGCAACTTCCA 781
Db 376 TrpThrGlyAspGluProLeuProTyGlnProThrPheSerAspAspTrpGlnLeuPro 395
QY 782 GAGCCCTCAGCAACGACCCCTTAGATACAGGACCCCTGTTCCCTTCGGCGGGAACT 841
Db 396 GluProSerSerGlnAlaProLeuGlyTyGlnAspProValSerLeu----- 411
QY 842 CATAGTTAGGGAGCACCTCACACTTTCTCAGGAGAAGACACACACCAACCATGACCTCAG 901
Db 411 ----- 411
QY 902 CTGGGGACCCCAACTCCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGAAATG 961
Db 412 -----ArgProSerSerProLysProAlaGlnProAlaLeuLysMet 425
QY 962 CAAGTCTTGTACGAGTTTGAAGCTAGGAACCCACGCGGAAGTACTGTGTCTCAGGAGAG 1021
Db 426 GlnValLeuTyGlnPheGluAlaArgAsnProArgGluLeuThrValGlnGlyGlu 445
QY 1022 AAGCTGGAGGTCTGGACACACAGCAGCGGTGGTGGCTGGTGAAGAAATCAGCGGGACGG 1081
Db 446 LysLeuGluValLeuAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArg 465
QY 1082 AGCGGTACATTCCAAGCAACATCTCTGGAGCCCTACAGCCGGGGACCCCTCGGACCCAG 1141
Db 466 SerGlyTyIleProSerAsnIleLeuGluProLeuGlnProGlyThrProGlyThrGln 485
QY 1142 GCCAGTCAACCTCTCGGCTTCCAATGCTTCGACTTAGCTAGCTCGAGCCCTCAAGAGGTACA 1201
Db 486 GlyGlnSerProSerArgValProMetLeuArgLeuSerSerArgProGluGluValThr 505
QY 1202 GACTGGCTCAGGACGAGAACTTCTCCACTGGCCAGCGGTGAGACACTTGGGTCCTGAGC 1261
Db 506 AspTrpLeuGlnAlaGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThr 525
QY 1262 GGGAGCCAGCTACTTCGCATAGACCTGGGGAGCTACAGATGCTATGTCCACAGAGGCC 1321
Db 526 GlySerGlnLeuLeuArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAla 545
QY 1322 CCACGAATCTCTCCCGCTGAGGCTGTCCAGAGGATCTGGGGATAGCCCT 1375
Db 546 ProArgIleLeuSerArgLeuGluAlaValArgArgMetLeuGlyIleSerPro 563
```

```
RESULT 4
Q91WL0
ID Q91WL0 PRELIMINARY; PRT; 600 AA.
AC Q91WL0;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to hypothetical protein FLJ21522 (Epidermal growth factor
DE receptor pathway substrate 8 related protein 3).
GN AI504489 OR EPS8R3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON;
RA Strausberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6;
RA Scita G., Confalonieri S., Offenhauser N., Borgonovo A., Tocchetti A.,
RA Romano P., Di Fiore P.P.;
RT "Cloning and characterization of novel members of the Eps8 protein
RT family";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC014734; AAL14734.1; -.
DR MGD; AY074932; AAL76121.1; -.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW Hypothetical protein; Receptor.
SQ SEQUENCE 600 AA; 68215 MW; 4AD0F34FCCCCDD39 CRC64;

Alignment Scores:
Pred. NO.: 1,38e-119 Length: 600
Score: 1606.50 Matches: 318
Percent Similarity: 78.97% Conservative: 50
Best Local Similarity: 68.24% Mismatches: 87
Query Match: 51.38% Indels: 11
DB: 11 Gaps: 6

US-09-762-021a-1 (1-1710) x Q91WL0 (1-600)
QY 2 GCAGAGCGACTGAAGACACGCTCGAGAGGCTCTGGAGAAAGAGCTGGAGCAAGACCT 61
Db 136 AlaGluGlnLeuArgThrSerLeuGlnLysAlaLeuGluGluGluLeuGluArgPro 155
QY 62 CCACCTGGAGGCTTCAGCCAGCCAGGACATGATGAGGGGCGCTGTATGGNAAGCGCG 121
Db 156 ArgPheGlyValIleHisProSerGlnAspArgTrpLysGlyProProLeuGluArgPro 175
QY 122 CTCCTCTATGAGCAGGACGCTATCTGAGCGC-----GGGATCCCT 163
Db 176 LeuProIleGlnGlnAlaProProLeuGluGlnArgPheSerProGluHisArgPhePro 195
QY 164 CCAGAACAGCCCAAGAGGACCCCTAGACACAGCCTCCCAACCATCCCAAGGCCCTTG 223
Db 196 ProGluGlnProHisAsnMetThrSerGluArgSerIleSerProSerArgSerLeu 215
QY 224 CCAGCCCAAGGACGCGCCGAGAACCAAGTGCCTTTACTCTG---CCTCTCCAAGCGCG 280
Db 216 ThrHisTyProSerAlaArgGluProAsnGlyPheThrLeuProProProArgArg 235
QY 281 TCCTCTTCCCGGAGGACCCAGAGGAGGACGAGGAGTGTCTGAACCATGCTCTAAGGGAC 340
Db 236 AlaProSerProGluAspProGluArgAspGluGluValLeuAsnHisValLeuArgAsp 255
QY 341 ATTAGCTCTTTCATGGGAAAGCTGGAGAAAGCGCCAGGCAAGACCAAGAGAGAG--- 397
```

```
Db 256 ileGluLeuPheAlaGlyLysLeuLysGluValGlnAlaArgAsnSerHisLysLysThr 275
Qy 398 MAATTGGG---AAAAAACAACAGCAGGAGGTCTCACCCAGGACAGTACATTGAC 454
Db 276 LysLeuGlyArgLysLysSerLysAsnGlyIleThrGlnAlaGluTrpIleAsp 295
Qy 455 TGTTCACAGAGATCAAGTACAGCTCAACCTCTCGGAAGGCTGGCCACCTGGCTGAAG 514
Db 296 CysPheGlnLysIleLysLysSerPheAsnLeuLeuGlyLysLeuAlaLeuArgMetGln 315
Qy 515 GAGCAAGTGCCTCGTACACATCTCTTCAGTCCCTGCACTTCACTGCTGCC 574
Db 316 GluThrSerAlaProGluPheValGlyLeuIlePheGlnThrLeuLysPheIleLeuSer 335
Qy 575 AGGTCCCTGAGGCTGAGCAGCCCAAGTGATCTCACCCCTCTCACCCCTCAAACT 634
Db 336 GlnCysProGluAlaGlyLeuProAlaLysValIleSerProLeuLeuThrProLysAla 355
Qy 635 ATCAACCTGCTACAGTCTGTCTAAGCCACCTGAGAGTAACCTTTGGATGGGTTGGC 694
Db 356 IleAspLeuLeuGlnSerCysLeuSerProProGluAspThrLeuTrpLysSerLeuGly 375
Qy 695 CCAGCTGGACCTAGCGGGCCGACTGGAGCGGATGAGCCCTGCCCTACCAACCC 754
Db 376 ThrSerTrpThrThrSerTrpAlaAspTrpThrGlySerGluProProProTrpGlnPro 395
Qy 755 ACATTCCTCAGATGCTGCACTTCCAGAGCCCTCCAGCCAGCACCCTTAGGATACAG 814
Db 396 ThrPheTrpAspGlyTrpGlnIleProGlnProArgSerMetMetProIleThrAsnGln 415
Qy 815 GACCCTGTTTCCCTCGCGGGGAAGTCATAGTTAGGAGCACCTCACACTTTCTCTCAG 874
Db 416 AspSerIleSerLeu---ArgGlySer---ArgMetArgSerSerLeuHisPheProArg 433
Qy 875 GAGAAGACACACACCACTGACCTCAGCTGGGGAGCCCACTCCAGCCCTCCAGCCCC 934
Db 434 AspGluProTrpAsnHisAsnProGluTrpGluAspSerAsnLeuProLeuSerSerPro 453
Qy 935 AAACCTGCCAGCCCTGAAATGCAAGTCTGTGACGAGTTTGAAGCTAGGAACCCA 994
Db 454 SerProGlyArgAlaAlaLeuLysMetGlnValLeuTrpGluPheGluAlaArgAsnAla 473
Qy 995 CGGGAACCTGCTGGTCCAGGAGAGAGCTGGAGGTTCTGGACCAAGCGGTGG 1054
Db 474 GlnGluLeuThrValAlaGlnGlyIleLeuGluValLeuAspGlnSerLysArgTrp 493
Qy 1055 TGGCTGGTCAAGATGAGCGGGGAGCGGAGGCTACATTCACAGCAACATCTGGAGCCC 1114
Db 494 TrpLeuValLysAsnGluAlaGlyLeuThrGlyTrpIleProSerAsnIleLeuGluPro 513
Qy 1115 CTACAGCGGGGACCCCTGGGACCCAGGCGCAGTCACCCCTCTCGGTTTCAATGCTCGA 1174
Db 514 LeuProAlaGlyAlaProArgGlyHisArgGlnProSerPheArgAlaProMetLeuArg 533
Qy 1175 CTTAGCTCAGGCGCTGAAGAGTTCACAGCTGGCTGCGAGGAGAGAACTTCTCCACTGCC 1234
Db 534 LeuSerSerLysProGluGluValThrAlaTrpLeuGlnAlaGluAsnPheSerThrVal 553
Qy 1235 ACGGTGAGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATTAAGACCTGGGAG 1294
Db 554 ThrValArgThrLeuGlySerLeuMetGlySerGlnLeuLeuHisMetArgProGlyGlu 573
Qy 1295 CTACAGATGCTATGCCACAGGAGCCCAAGTCTCTGTCGCCGCTGAGGCTCTCAGA 1354
Db 574 LeuGlnMetLeuLysProGlnGluAlaProArgIleGlnAlaArgLeuAspAlaValArg 593
Qy 1355 AGGATGCTGGGATAAGC 1372
Db 594 ArgMetLeuGlyMetThr 599
RESULT 5
Q99K30
```

```
ID Q99K30 PRELIMINARY; PRT; 729 AA.
AC Q99K30;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Similar to hypothetical protein FLJ21935.
GN A1042819.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; BC005492; AA005492.1; -.
DR HSP; Q08509; IAOJ.
DR MGI; MGI:2138828; A1042819.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00462; PTB; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS01179; PID; 1.
DR PROSITE; PS00002; SH3; 1.
DR Hypothetical protein; SH3 domain.
KW SEQUENCE 729 AA; 82229 MW; A11DE82FF2C0BD18 CRC64;
SQ
Alignment Scores:
Pred. No.: 3,7e-39 Length: 729
Score: 597.50 Matches: 173
Percent Similarity: 46.21% Conservative: 95
Best Local Similarity: 29.83% Mismatches: 189
Query Match: 19.11% Indels: 124
DB: 22
US-09-762-021A-1 (1-1710) x Q99K30 (1-729)
Qy 5 GAGCGACTGAGACAGCAGCTCGAGAGGCTCTGGAGGAGAGCTGGACAAAGACTCGA 64
Db 178 LysLysMetArgProGlnThrLeuLysGlyHisGlnGluLysIleArgGlnArgGlnSer 197
Qy 65 CTGGAGGCTTCAGCCAGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 124
Db 198 Ile-----LeuProProProGln-----SerProAla-----ProIle 208
Qy 125 CCTATGGAGCAGGACGCTATCTGGAGCGGGGATCTCCAGAACAGCCCCACAGAGG 184
Db 209 ProPheGlnArg-----GlnProGlyAspSerProGlnAlaLysAsnArgVal 224
Qy 185 ACCCTAGAGCAGCCTCCACCATCCCAAGGCCCTGCCAGCCGACACAGTGCCTCGA 244
Db 225 GlyLeu-----ProLeuProValPro----- 231
Qy 245 GAACCAAGTGCCTTTACTCTGCTCTCT--CCAAGGCGGTCTCTTCCCGAGGACCCA 301
Db 232 -----PheSerGluProGlyTrpArgArgGluSerGlnAspGluGluPro 247
Qy 302 -----GAGAGGACGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 337
Db 248 ArgAlaValLeuAlaGlnArgIleGluLysGluThrGlnIleLeuAsnCysThrLeuAsp 267
Qy 338 GACATTGAGCTTTCATGGGAAAGCTGGAGAGGCGCCAG-----GCAAGACACAGC 388
Db 268 AspIleGluTrpPheValAlaArgLeuGlnLysAlaAlaGluAlaPheLysGlnLeuAsn 287
Qy 389 AGGAAGAGAAATTTGGGAAAAAACAAGAC-----CAGGAGGCTCTACCC--- 436
Db 288 GlnArgLysLysGlyLysLysAsnLysLysGlyProAlaGluGluValIleThrLeu 307
Qy 437 -----CAGGCACAGTACATTGACTGCTTCCAGAGATCAAGTACAGC 478
```

```
Db 308 ArgAlaArgProProSerGluGlyGluPheValAspCysPheGlnLysThrLysLeuAla 327
      ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY 479 TTCAACCTCTGGAGGCTGGCCACCTGGCTCAAGGACAGAGTCCCTGAGCTCGTA 538
      ||||| ||||| ::::: ::::: ::::: ||||| ||||| |||||
Db 328 IleAsnLeuAlaLysLeuGlnLysHisIleGlnAsnProSerAlaAlaGluLeuVal 347
      ||||| ||||| ::::: ::::: ::::: ||||| ||||| |||||
QY 539 CACATCTCTTCAAGTCCCTGAATTCATCTCGGCGGAGTGGCCTGAGGCTGGCCTAGCA 598
      ||||| ||||| ::::: ::::: ||||| ||||| |||||
Db 348 HisPheLeuPheGlyProLeuAspLeuIleIleAsnThrCysGlySerProAspIleAla 367
      ||||| ||||| ::::: ::::: ||||| ||||| |||||
QY 599 GCCCAAGTATCTACCCCTCCACCTCAAGCTATCAAGCTGTACAGTCTCTGTCTA 658
      ||||| ||||| ||||| ||||| ::::: ||||| |||||
Db 368 ArgSerValSerSerProLeuLeuSerThrAspAlaValSerPheLeuArgGlyHisLeu 387
      ||||| ||||| ||||| ||||| ::::: ||||| |||||
QY 659 AGCCCACTGAGAGTAACCTTTGGATGGGTGGGCGGAGCTGGACCACTAGCCGGCC 718
      ||||| ||||| ||||| ||||| ||||| |||||
Db 388 ValProLysGluMetThrLeuThrGluSerLeuGlyGluThrTrpMetArgProArgSer 407
      ||||| ||||| ||||| ||||| ||||| |||||
QY 719 GACTGGACAGGCGATGAGCCCTGGCC---TACCAACCCACATTCACAGATGACTGGCAA 775
      ::::: ::::: ||||| ||||| ||||| |||||
Db 408 GluTrpProArgGluProGlnValProLeuTyrValProLysPheArgSerGlyTrpGlu 427
      ||||| ||||| ||||| ||||| ||||| |||||
QY 776 CTTTCCA---GAGCCCTCCAGCCAGCACCCCTTAGGATACCAAGAC-----CCT 820
      ||||| ::::: ||||| ||||| ||||| |||||
Db 428 ProProLeuAspValLeuGlnGluAlaProTrpGluValGluGlyLeuAlaSerValPro 447
      ||||| ||||| ||||| ||||| ||||| |||||
QY 821 GTTTCCTTCGGCGGGAAGTCACTAGTTAGGAGCACCTCACACTTCTCAGGAGAAG 880
      ||||| ||||| ||||| ||||| ||||| |||||
Db 448 SerAspGlnLeuThrProLysAsnArgLeu---SerValArgHisSerProLysHisSer 466
      ||||| ||||| ||||| ||||| ||||| |||||
QY 881 ACACACAACCATGACCTCAGCCTGGGGAC-----CCCAACTCCAGGCCCTCCAGC 931
      ::::: ||||| ||||| ||||| ||||| |||||
Db 467 LeuSerSerGluSerGlnAlaProGluAspIleAlaProGlySerSerProHisAla 486
      ||||| ||||| ||||| ||||| ||||| |||||
QY 932 CCCAAACCTGCCAGCCAGCCCTGAAATG-----CAAGCTGTGACGATTT 979
      ::::: ||||| ||||| ||||| ||||| |||||
Db 487 AsnArgGlyTyrGlnProThrProAlaMetThrLysTyrValLysIleLeuTyrAspPhe 506
      ||||| ||||| ||||| ||||| ||||| |||||
QY 980 GAAGCTAGGAAGCCCGGAACTGACTGTGGTCCAGGAGAGAGCTGGAGTCTGGAC 1039
      ||||| ||||| ||||| ||||| ||||| |||||
Db 507 ThrAlaArgAsnAlaAsnGluLeuSerValLeuLysAspGluValLeuGluValLeu 526
      ||||| ||||| ||||| ||||| ||||| |||||
QY 1040 CACAGCAAGCGTGGTGGTGAAGATGAGCGGGGAGCGGCTACATTCACAAGC 1099
      ::::: ||||| ||||| ||||| ||||| |||||
Db 527 AspGlyArgGlnTrpTrpLysLeuArgAsnArgSerGlyGlnAlaGlyTyrValProCys 546
      ||||| ||||| ||||| ||||| ||||| |||||
QY 1100 AACATCTG-----GAGCCCTACAG----- 1120
      ||||| ||||| ||||| ||||| ||||| |||||
Db 547 AsnIleLeuAlaGluAlaArgGlnGluAspValGlyAlaProLeuGluGlnSerGlyGln 566
      ||||| ||||| ||||| ||||| ||||| |||||
QY 1121 -----CCGGGGACCCCT----- 1132
      ||||| ||||| ||||| ||||| ||||| |||||
Db 567 LysTyrTrpGlyProAlaSerProThrHisLysLeuProProIlePheAlaGlyAsnLys 586
      ||||| ||||| ||||| ||||| ||||| |||||
QY 1132 ----- 1132
      ||||| ||||| ||||| ||||| ||||| |||||
Db 587 GluGluLeuIleHisMetAspGluValAsnAspGluLeuMetLysLysIleSerHis 606
      ||||| ||||| ||||| ||||| ||||| |||||
QY 1133 ---GGGACCCAGCCAGTCACTCCCTCTCGGGT-----CCAATG----- 1168
      ||||| ||||| ||||| ||||| ||||| |||||
Db 607 IleLysThrGlnProGlnArgAsnPheArgValGluArgSerGlnProValHisLeuPro 626
      ||||| ||||| ||||| ||||| ||||| |||||
QY 1169 CTTTCGACTTAGCTCAGGCGCTGAAGAGTCCACAGACTGGCTGCAGGCAGAGAATTCTCC 1228
      ||||| ||||| ||||| ||||| ||||| |||||
Db 627 LeuThrPheGluSerGlyProAspGluValArgAlaTrpLeuGluAlaLysAlaPheSer 646
      ||||| ||||| ||||| ||||| ||||| |||||
QY 1229 ACTGCCAGGTGAGACACTTGGGTCCCTCAGCGGGAGCGACGACTTGCATTAAGACCT 1288
      ||||| ||||| ||||| ||||| ||||| |||||
Db 647 AlaArgIleValGluAsnLeuGlyIleLeuThrGlyProGlnLeuPheSerLeuAsnLys 666
      ||||| ||||| ||||| ||||| ||||| |||||
QY 1289 GGGGAGCTACAGATGTATGTCCACAGGAGGCCCCACGAATCTGTCCCGCTGGAGGCT 1348
      ||||| ||||| ||||| ||||| ||||| |||||
```

```
Db 667 GluGluLeuLysLysValCysGlyGluGlySerArgValTyrSerGlnLeuThrVal 686
QY 1349 GTCAGAAGGATCGCTGGG-CATAAGCCCTTAGCCACACGCTTAGACACCTCCAAGAACCAG 1407
      ::::: ||||| ::::: ||||| ||||| ||||| |||||
Db 687 GlnLysAlaPheLeuGluLysGlnSerGlySerGluLeuGluLysLeuMetSerLys 706
      ||||| ||||| ||||| ||||| ||||| |||||
QY 1408 GCCCCCGCTGATCAAGATGGCAGATCTGATACCCATTAGAGCCCGGAGAAATTCCTTCT 1467
      ::::: ||||| ::::: ||||| ||||| |||||
Db 707 IleArgArgAlaGluAspSerTyrThrSerGlnHisThrSerProGluSerGluGlyAla 726
      ||||| ||||| ||||| ||||| ||||| |||||
RESULT 6
QYH6S3
AC Q9H6S3 PRELIMINARY; PRT; 743 AA.
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DE 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE CDNA: FLJ21935 fis, clone HEP04373.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Ota T., Suzuki Y., Obayashi M., Nishi T., Shibahara T.,
RA Tanaka T., Nakamura Y., Isogai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AK025588; BAB15180.1; -.
DR HSSP; Q08509; 1A0J.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00462; PTB; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS01179; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 743 AA; 83792 MW; DB01D8F6363A1F80 CRC64;
```

Alignment Scores:

Pred. No.:	2,12e-38	Length:	743
Score:	568.00	Matches:	162
Percent Similarity:	44.72%	Conservative:	71
Best Local Similarity:	31.09%	Mismatches:	156
Query Match:	18.80%	Indels:	132
DB:	4	Gaps:	18

US-09-762-021A-1 (1-1710) x Q9H6S3 (1-743)

```
QY 131 GAGCAGGACGCTATCTGGAGCGGGGATCCCTCCAGAACAGCCCCACCAGAGGACCCTA 190
      ||||| ||||| ::::: |||||
Db 218 GluLysIleArgGlnArgGlnSerPheLeuProPro----- 229
      ||||| ||||| ||||| ||||| ||||| |||||
QY 191 GAGCAGAGCCTCCCAACCATCCCAAGGCCCTGCCA---CGCCACACCACTGCCCCAGAA 247
      ||||| ||||| ||||| ||||| ||||| |||||
Db 230 -----ProGlnGlyProAlaProIleProPheGlnHisArgGlyGlyAspSer 245
      ||||| ||||| ||||| ||||| ||||| |||||
QY 248 CCAAGTGCC-----TTTACTCTGCCTCCTCCCAAGCGCG 280
      ||||| ||||| ||||| ||||| ||||| |||||
Db 246 ProGluAlaLysAsnArgValGlyProGlnValProLeuSerGluProGlyPheArgArg 265
      ||||| ||||| ||||| ||||| ||||| |||||
QY 281 TCCTCTTCCCGAGACCCCA-----GAGAGGGACGAGGAA 316
      ||||| ||||| ||||| ||||| ||||| |||||
Db 266 ArgGluSerGlnGluProArgAlaValLeuAlaGlnLysIleGluLysGluThrGln 285
      ||||| ||||| ||||| ||||| ||||| |||||
QY 317 GTGCTCAACCATCTCCCTAAGGAGCATTTGAGCTGTTTCATGGAAAGCTGGAGAGGCC--- 373
      ||||| ||||| ||||| ||||| ||||| |||||
Db 286 IleLeuAsnCysAlaLeuAspAspIleGluTrpPheValAlaArgLeuGlnLysAlaAla 305
```

```
QY 374 -----CAGCAAGACAGCAGGAGAGAAATTTGGGAAAAA 412
Db 306 GluAlaPheLysGlnLeuAsnGlnArgLysLysGlyLysLys-----GlyLysLys 323
QY 413 ACAAGGACAGGAGGAGTCTAC-----CAGCACACATACATT 451
Db 324 AlaProAlaGluGlyValLeuThrLeuArgAlaArgProProSerGluGlyGluPheLe 343
QY 452 GACTGCTCCAGAGCAAGTCAACCTCTCCGGAAGGCTGCCACCTGGCTG 511
Db 344 AspCysPheGlnLysLeuAlaLeuLeuLeuAlaLysLeuGlnLysHisLe 363
QY 512 AAGGACAGAAGTCCCTGAGCTGCTGACACATCTCTCAAGTCCCTCAACTCATCTG 571
Db 364 GlnAsnProSerAlaAlaGluLeuValHisPheLeuGlyProLeuAspLeuLeVal 383
QY 572 GCCAGGTGCCGTGAGCTGAGCAGCCCAAGTGTCTACCCCTCTCACCCCTTAA 631
Db 384 AsnThrCysSerGlyProAspAlaAlaArgSerValSerCysProLeuLeuSerArgAsp 403
QY 632 GCTATCAACCTGCTACAGTCCCTCTAAAGCCACCTGAGAGTAACTTTGGATGGGTG 691
Db 404 AlaValAspPheLeuArgGlyHisLeuValProLysGluMetSerLeuTrpGluSerLeu 423
QY 692 GCCCAGCTGAGCAGCTAGCCGGCCGAGCTGGACAGCGGATGAGCCCTGCC---TAC 748
Db 424 GlyGluSerTrpMetArgProArgSerGluTrpProArgGluProGlnValProLeuTrp 443
QY 749 CAACCCACATCTCAGATGAGTGGCACTTCCA----- 781
Db 444 ValProLysPheHisSerGlyTrpGluProProValAspValLeuGlnGluAlaProTrp 463
QY 782 -----GAGCCCTCCAGCAGCAGCCCTTAGGATACCAGGACCCGTTTCCCTCGCGG 835
Db 464 GluValGluGlyLeuAlaSerAlaProLeuGluGluValSerProValSer-----Arg 481
QY 836 GGAAGTCATAGTTAGGAGGACCTTCACACTTTCTCAGGAGAGACACACAACCATGAC 895
Db 482 GlnSerIleArg---AsnSerGlnLysHisSerPro-----ThrSerGluProThr 497
QY 896 CTTAGCCTGGGAC-----CCCACTCCAGGCTTCAGCCCAACCTGCCAG 946
Db 498 Pro---ProGlyAspAlaLeuProProValSerProHisThrHisArgGlyTrpGln 516
QY 947 CCA-----GCCTGAAATGCAAGTCTGTACGAGTTTGAAGCTAGGAACCCA 994
Db 517 ProThrProAlaMetAlaLysTrpValLysIleLeuLeuTrpAspPheThrAlaArgAsnAla 536
QY 995 CGGGAAGTCTGCTGCTCCAGGAGAGAGCTGGAGGTTCTGACACAGCAGGCGGTG 1054
Db 537 AsnGluLeuSerValLeuLysAspGluValLeuGluValLeuGluAspGlyArgGlnTrp 556
QY 1055 TGGCTGGTGAAGATGAGCGGAGGAGGCTGCTACATTCNAGCAATCCTG----- 1108
Db 557 TrpLysLeuArgSerArgSerGlyGlnAlaGlyTrpValProCysAsnIleLeuGlyGlu 576
QY 1108 ----- 1108
Db 577 AlaArgProGluAspAlaGlyAlaProPheGluGlnAlaGlyLysTrpTrpGlyPro 596
QY 1109 -----GAGCCCTCCAGCGGGAGCCCTGGGACCCAGGCGCAG----- 1147
Db 597 AlaSerProThrHisLysLeuProProSerPheProGlyAsnLysAspGluLeuMetGln 616
QY 1147 ----- 1147
Db 617 ArgMetAspGluValAsnAspGluLeuIleArgLysIleSerAsnIleArgAlaGlnPro 636
QY 1148 -----TCACCTCTCGGTTCCAAATCTTCGACTTAGC 1180
Db 637 GlnArgHisPheArgValGluArgSerGlnProAlaSerGlnPro---LeuThrTrpGlu 655
QY 1181 TCGAGGCTGAAGAGTTCACAGACTGGCTGAGGAGGAGAACTTCTCCACTGCCACGGTG 1240
```

```
Db 656 SerGlyProAspGluValArgAlaTrpLeuGluAlaLysAlaPheSerProArgIleVal 675
QY 1241 AGGACACTGGCTCCTGACGGGAGCCAGCTACTTCGCATAGACCTGGGAGCTACAG 1300
Db 676 GluAsnLeuGlyIleLeuThrGlyProGlnLeuPheSerLeuAsnLysGluGluLeuLys 695
QY 1301 ATGCTATCTCCACAGAGGCCCCACGAATCTCTCCGCTGGAGCTGTCTCAGAAAGGATG 1360
Db 696 LysValCysGlyGluGluGlyValArgValTySerGlnLeuThrMetClnLysAlaPhe 715
QY 1361 CTG 1363
Db 716 Leu 716

RESULT 7
Q9H6K9 PRELIMINARY; PRT; 715 AA.
ID Q9H6K9 AC Q9H6K9;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE CDNA: FLJ22171 fis, clone HRC00654 (Epidermal growth factor receptor
DE pathway substrate 8 related protein 2).
GN EPS8R2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Ohayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isono T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Scita G., Confalonieri S., Offenhauser N., Borgonovo A., Tocchetti A.,
RA Romano P., Di Fiore P.P.;
RT "Cloning and characterization of novel members of the Eps8 protein
RT family.";
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
CC -! SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AK025824; BAB15248.1; -
DR EMBL; AY074929; AAL76118.1; -
DR HSSP; Q08509; IAOJ.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00482; PTH; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS01179; PID; 1.
DR PROSITE; PS00002; SH3; 1.
KW Receptor; SH3 domain.
SQ SEQUENCE 715 AA; 80620 MW; DAB07744B04CFEE2 CRC64;

Alignment Scores:
Pred. No.: 4.39e-38 Length: 715
Score: 584.00 Matches: 163
Percent Similarity: 45.00% Conservative: 71
Best Local Similarity: 31.35% Mismatches: 156
Query Match: 18.68% Indels: 130
DB: Gaps: 19

US-09-762-021A-1 (1-1710) x Q9H6K9 (1-715)
QY 131 GAGCAGGACGCTATCTATCGAGCCGGGATCCCTCCAGACAGCCCCACCAGGACCCTA 190
Db 190 GluLysIleArgGlnArgGlnSerIleLeuProPro----- 201
QY 191 GAGCAGACGCTCCCAACCATCCCAAGGCCCCCTGGCCA---CGCCACACCACTGCCCGAGAA 247
```



```

Qy 167 GAACAGCCCCACAGGACCTAGAGCACAGCCTC-----CCACCATTCCCAAGG 217
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
69 ValArgAlaValIleSerThrValGluArgGlyAlaGlyArgGlyProGlnAlaLys 88
Qy 218 CCCTGCCAGCCACACAGTCCGAGAACCAAGTGCCTTACTCTCCCTCCCAAGG 277
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
89 ProIleProGluAlaGluGlnArgProGluProValGlyThrSerSerAsnAla 108
Qy 278 CGGTCTCTTCCCGGAG-----GACCCAGAG 304
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
109 AspSerAlaSerProAspLeuGlyProArgGlyProAspLeuAlaValLeuGlnAlaGlu 128
Qy 305 AGGAGAGGAGTCTCTGAACCATCTCTAAGGGACATTGAGCTTTCATGGGAAGCTG 364
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
129 ArgGluValAspIleLeuAsnHisValPheAspValGluSerPheValSerArgLeu 148
Qy 365 GAGAAAGGCC-----CAGCGAAAGACCAGCAGGAGAGAATAATTGGGAAAAAACAAG 418
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
149 GlnLysSerAlaGluAlaAlaArgValLeuGluHisArgGluArgGlyArgSerArg 168
Qy 419 -----GACCCAGGAGGTCTCACC-----CAGGCACAG 445
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
169 ArgArgAlaAlaGlyGluGlyLeuLeuThrLeuArgAlaLysProProSerGluAlaGlu 188
Qy 446 TACATTGACTGCTCCAGAGATCAAGTACAGCTTCAACCTCTCTGGGAAGCTGGCCACC 505
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
189 TyrThrAspValLeuGlnLysIleLysTyrAlaPheSerLeuLeuAlaArgLeuGly 208
Qy 506 TGGCTGAGGACAGCAAGTGCCTGAGCTCGTACACATCTCTCAAGTCTCCTGAACATTC 565
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
209 AsnIleAlaAspProSerSerProGluLeuLeuHisPheLeuPheGlyProLeuGlnMet 228
Qy 566 ATCTCTGGCCAGGTGGCTGGCTGAGCGCTAGCAGCCCAAGTGATCTCACCCTCTCACC 625
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
229 IleValAsnThrSerGlyGlyProGluPheAlaSerSerValArgArgProHisLeuThr 248
Qy 626 CTTAAAGTATCAACCTCTACAGTCTCTTAAGCCCACTGAGAGTAACCTTTGGATG 685
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
249 SerAspAlaValAlaLeuLeuArgAspAsnValThrProArgGluAsnGluLeuTrpThr 268
Qy 686 GGGTTGGCCAGCTGACCACTAGCCGGCGGCTGAGCGGCTGAGCGGCTGAGCGGCTGCC 745
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
269 SerLeuGlyAspSerTrpThrArgProGlyLeuGluLeuSerProGluGlyProPro 288
Qy 746 TACCAACCCACATTCATGACTGGCACTTCA-----GAGCCCTCCAGCAAGCA 799
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
289 TyrArgProGluPhePheSerGlyTrpGluProProValThrAspProGlnSerArgAla 308
Qy 800 CCTTTAGGATACAGGACCCCTGTTTCTTCCCTCGCGGGGAAATCATAGTTAGGGAGCAC 859
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
309 -----TrpGluAspProValGluLysGlnLeuGlnHisGluArgArgArgGln 325
Qy 860 TCACACTTTCCTCAGGAGAACACACACACAT---GACCCCTCAGCTGGGACCCCAAC 916
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
326 GlnSerAlaProGlnValAlaValAsnGlyHisArgAspLeuGluPro----- 341
Qy 917 TCCAGGCTCTCAGGCCCAACCTGCCAGCCAGCCCTGAAATGCAAGTCTTG----- 970
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
342 -----GluSerGluProGlnLeuGluSerGluThrAlaGlyLysTrpValLeuCysAsn 359
Qy 971 TACGAGTTTGAAGCTAGGAACCCAGCGGAATGACTGTGGTCCAGGGAGAGAGCTGGAG 1030
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
360 TyrAspPheGlnAlaArgAsnSerSerGluLeuSerValLysGlnArgAspValLeuGlu 379
Qy 1031 GTTCTGGACACAGACCGGTGGTGGTGGTGAAGATGAGGGGACGAGCGGCTAC 1090
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
380 ValLeuAspSerArgLysTrpTrpLysValArgAspProAlaGlyGlnGluGlyTyr 399
Qy 1091 ATTCCAAGCACATCTTGAGCCCTACAGCGCGGG----- 1126
Db   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
400 ValProTyrAsnIleLeuThrPro---TyrProGlyProArgLeuHisSerGlnSer 418
Qy 1127 -----ACCCCTGGG----- 1135

```

```

Db 419 ProAlaArgSerLeuAsnSerThrProProProProAlaProAlaProAlaPro 438
Qy 1135 ----- 1135
Db 439 ProAlaLeuAlaArgProArgTrpAspArgProArgTrpAspSerCysAspSerLeuAsn 458
Qy 1135 ----- 1135
Db 459 GlyLeuAspProSerGluLysGluLysPheSerGlnMetLeuIleValAsnGluGluLeu 478
Qy 1136 -----ACCCAGGCGCAGTCA---CCCTCTCGGGTT----- 1162
Db 479 GlnAlaArgLeuAlaGlnGlyArgSerGlyProSerArgAlaValProGlyProArgAla 498
Qy 1163 -----CCAATGCTTCGACTTAGCTCGAGGCTGAAGAGGTCACAGACTGGCTGCAGGCA 1216
Db 499 ProGluProGlnLeuSerProGlySerAspAlaSerGluValArgAlaTrpLeuGlnAla 518
Qy 1217 GAGAACTTCTCCACTGCCAGCTGAGGACACTTGGGTCCCTCAGCGGAGGAGCAGTACTT 1276
Db 519 LysGlyPheSerSerGlyThrValAspAlaLeuGlyValLeuThrGlyAlaGlnLeuPhe 538
Qy 1277 CGCATAAGACCTGGGGAGCTACAGATGCTATGTCACAGGAGGCCCCAGAACTCTGTCC 1336
Db 539 SerLeuGlnLysGluGluLeuArgAlaValSerProGluGluGlyAlaArgValTyrSer 558
Qy 1337 CGGCTGGAGGCTGTGCAAGAGTACTG 1363
Db 559 GlnValThrValGlnArgSerLeuLeu 567

RESULT 9
Q8TE68 PRELIMINARY; PRT: 723 AA.
AC Q8TE68;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Epidermal growth factor receptor pathway substrate 8 related protein
DE 1
GN EPS8R1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Scita G., Confalonieri S., Offenhauser N., Borgonovo A., Tocchetti A.,
RA Romano P., Di Fiore P.P.;
RT "Cloning and characterization of novel members of the Eps8 protein
RT family."
RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY074928; AAL76117.1; -.
KW Receptor.
SQ SEQUENCE 723 AA; 80300 MW; 00BCDDFE16FG2E13 CRC64;

Alignment Scores:
Pred. No.: 3,63e-37 Length: 723
Score: 572.50 Matches: 168
Percent Similarity: 44.29% Conservative: 84
Best Local Similarity: 29.53% Mismatches: 184
Query Match: 18.31% Indels: 133
DB: Gaps: 19

US-09-762-021a-1 (1-1710) x Q8TE68 (1-723)
Qy 2 GCAGAGCGACTGAAGACCCAGCTGCAGAAAGCTCTGGAGAGAGCTGGAGCAAGA--- 58
Db 144 AlaGluLeuIleArgGluAspIleGlnGlyAlaLeuHisAsnTyrArgSerGlyArgGly 163
Qy 59 ---CCTCAGCTGGAGGCTTCAGCCAGCCAGCCAGGAC-----AGATGAGGGGGCCT 106
Db 164 GluArgArgAlaAlaLeuArgAlaThrGlnGluGluLeuGlnArgAspSerPro 183

```

```
QY 107 GCTATGAAGGCCGCTCCCTATGAGCAGCAGCAGCTATCTGGAGCGGGATCCCTCCA 166
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 184 AlaAlaGluThrPro--ProLeuGlnArgArg-----ProSer 195

QY 167 GAACAGCCCAACAGGACCCCTAGACACAGCTC-----CCACATCCCAAGG 217
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 196 ValArgAlaValIleSerThrValGluArgGlyAlaGlyArgGlyProGlnAlaLys 215

QY 218 CCCTCGCCAGCCACACACAGCTGCGCAGAACCAAGTGCCTTTACTCTGCTCCTCAAGG 277
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 216 ProIleProGluAlaGluAlaGlnArgProGluProValGlyThrSerSerAla 235

QY 278 CGCTCCTCTTCCCGGAG-----GACCAGCAG 304
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 236 AspSerAlaSerProAspLeuGlyProArgGlyProAspLeuAlaValLeuGlnAlaGlu 255

QY 305 AGGAGCAGAGTCTGAACCATGCTTAAGGACATTGAGCTGTTCATGGGAAGCTG 364
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 256 ArgGluValAspIleLeuAsnHisValPheAspAspValGluSerPheValSerArgLeu 275

QY 365 GAGAAGGCC-----CAGCAAGACACAGCAGGAAGAATAATTTGGGAAAAAACAAG 418
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 276 GlnLysSerAlaGluAlaAlaArgValLeuGluHisArgGluArgGlyArgSerArg 295

QY 419 -----GACCAGGAGGTCTCAC-----CAGGCACAG 445
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 296 ArgArgAlaAlaGlyGluGlyLeuLeuThrLeuArgAlaLysProProSerGluAlaGlu 315

QY 446 TACATTGACTGCTTCAGAAAGATCAAGTACAGCTTCAAGTCTCTGGAAGGCTGGCCACC 505
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 316 TyrThrAspValLeuGlnLysIleLysTyrAlaPheSerLeuLeuAlaArgLeuArgGly 335

QY 506 TGGCTCAAGGAGACAAGTCCCTGAGCTCGTACACATCTCTCAAGTCCCTGAACCTTC 565
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 336 AsnIleAlaAspProSerProGluLeuLeuHisPheLeuPheGlyProLeuGlnMet 355

QY 566 ATCTGCGCCAGGTGCCTGAGGTGCGCTAGCCCAAGTATCTCACCCCTCTCTCAC 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 356 IleValAsnThrSerGlyGlyProGluPheAlaSerSerValArgArgProHisLeuThr 375

QY 626 CCTAAGCTATCAACCTGCTACAGTCTCTTAAGCCCACTGAGAGTAACTTTGGATG 685
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 376 SerAspAlaValAlaLeuLeuArgAspAsnValThrProArgGluAsnGluLeuTrpThr 395

QY 686 GGGTGGCCGCGCCGCTGACCTAGCCGCGGCGGACTGGACAGGCGATGAGCCCTGCC 745
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 396 SerLeuGlyAspSerTrpThrArgProGlyLeuGluLeuSerProGluGlyProPro 415

QY 746 TACCAACCCACATCTCAGTGAAGTTCCTCA-----GAGCCCTCCAGCCCAAGCA 799
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 416 TyrArgProGluPhePheSerGlyTrpGluProProValThrAspProGlnSerArgAla 435

QY 800 CCCTTAGGATACAGACCCCTGTTTCCCTTGGCGGGGAGTATAGTTAGGAGCACC 859
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 436 -----TrpGluAspProValGluLysGlnLeuGlnHisLysLysArgArgGln 452

QY 860 TCACACTTTCTCAGAGACACACACACCAT--GACCCCTCAGCTGGGACCCCAAC 916
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 453 GlnSerAlaProGluValAlaValAsnGlyHisArgAspLeuGluPro----- 468

QY 917 TCCAGGCCCTCAGCCCAACCTGCGCCAGCCAGCCCTGAAATGCAAGTCTTG----- 970
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 469 -----GluSerGluProGlnLeuGluSerGluThrAlaGlyLysTrpValLeuCysAsn 486

QY 971 TAGGAGTTTGAAGTACAGCAACCCAGCACTGCTGGTCCAGGAGAGAGAGCTGGAG 1030
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 487 TyrAspPheGlnAlaArgAsnSerSerGluLeuSerValLysGlnArgAspValLeuGlu 506

QY 1031 GTTCTGACACAGCAGCGGTGGTGGTCAAGAAATGAGCGGCGGACGCGCTAC 1090
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 507 ValLeuAspAspSerArgLysTrpTrpLysValArgAspProAlaGlyGlnGlyTyr 526
```

```
QY 1091 ATTCCAAGCAACATCCTGGAGCCCTACAGCCGGG----- 1126
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 527 ValProTyrAsnIleLeuThrPro---TyrProGlyProArgLeuHisSerGlnSer 545

QY 1127 -----ACCCTGGG----- 1135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 546 ProAlaArgSerLeuAsnSerThrProSerProProProAlaProAlaProPro 565

QY 1135 ----- 1135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 566 ProAlaLeuAlaArgProArgTrpAspArgProArgTrpAspSerCysAspSerLeuAsn 585

QY 1135 ----- 1135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 586 GlyLeuAspProSerGluLysGluLysPheSerGlnMetLeuIleValAsnGluGluLeu 605

QY 1136 -----ACCAGGGCCAGTCA---CCCTCGGGT----- 1162
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 606 GlnAlaArgLeuAlaGlnGlyArgSerGlyProSerArgAlaValProGlyProArgAla 625

QY 1163 -----CCAATGCTTCGACTTAGCTGAGGCGCTGAAGAGGTACAGACTGGCTGCAGGCA 1216
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 626 ProGluProGlnLeuSerProGlySerAspAlaSerGluValArgAlaTrpLeuGlnAla 645

QY 1217 GAGAACTTCTCCACTCCACGCTGAGGACACTTGGCTCCCTGACGGGAGCCAGCTACT 1276
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 646 LysGlyPheSerSerGlyThrValAspAlaLeuGlyValLeuThrGlyAlaGlnPhe 665

QY 1277 CCATTAAGACCTGGGAGCTACAGATGCTATGTCCACAGGAGGCCCCCAGAACTCTGTC 1336
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 666 SerLeuGlnArgGluGluLeuArgAlaValSerProGluGluGlyAlaArgValTyrSer 685

QY 1337 CGCTGAGGCTGTCAAGAGGATGCTG 1363
   : : : : : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 686 GlnValThrValGlnArgSerLeuLeu 694

RESULT 10
Q9NXHO
ID Q9NXHO PRELIMINARY; PRT; 596 AA.
AC Q9NXHO:
DT 01-OCT-2000 (TRENBLrel. 15, Created)
DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE CDNA FLJ20258 fis, clone COLF7250.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=COLON MUCOSA;
RA Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA Suzuki Y., Umayashi M., Nishi T., Shibahara T., Tanaka T.,
RA Nakamura Y., Isoigai T., Sugano S.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -!- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AK00265; BAA91041.1; -.
DR HSP; Q08509; IAOJ.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 596 AA; 66517 MW; 3A586087464F7812 CRC64;

Alignment Scores:
Pred. No.: 4, 21e-37 Length: 596
Score: 571.50 Matches: 169
Percent Similarity: 44.23% Conservative: 83
Best Local Similarity: 29.70% Mismatches: 184
Query Match: 18.28% Indels: 133
```


DB: 4 Gaps: 19
US-09-762-021a-1 (1-1710) x QNKH0 (1-596)
QY 2 GCAGAGCCACTGAACACGAGCTCGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGA---- 58
||||| : : : : : |||||
Db 17 AlaGluLeuIleArgGluAspIleGlnGlyAlaLeuHisAsnTyrArgSerGlyArgGly 36
QY 59 ---CCTGACTTGGAGGCTTCAGCCAGCCAGGAC-----AGATGGAGGGGCT 106
||||| : : : : : |||||
Db 37 GluArgArgAlaAlaLeuArgAlaThrGlnGluLeuGlnArgAspArgSerPro 56
QY 107 GCTATGGAAGCCGCTCCCTATGAGAGCAGCAGCTATCTGGAGCCGGGATCCCTCCA 166
: : : ||| : : : : : : : : : :
Db 57 AlaAlaGluThrPro---ProLeuGlnArg-----ProSer 68
QY 167 GAACAGCCCCCAGCAGCAGCTAGAGCAGCAGCTC-----CCACCATCCCAAGG 217
: : : : : ||||| : : : : :
Db 69 ValArgAlaValIleSerThrValGluArgGlyAlaGlyArgGlyArgProGlnAlaLys 88
QY 218 CCCCTGCCAGCCACACAGCTGCCGAGAACCAAGTGCCTTACTCTGCCTCCTCCCAAGG 277
: : : : : ||||| : : : : :
Db 89 ProIleProGluAlaGluGlnAlaGlnArgProGluProValGlyThrSerSerAsnAla 108
QY 278 CGTCTCTCTCCCGGAG-----GACCCAGAG 304
||||| : : : : : |||||
Db 109 AspSerAlaSerProAspLeuGlyProArgGlyProAspLeuAlaValLeuGlnAlaGlu 128
QY 305 AGGAGCAGGAGTCTCAACCATGCTCTAAGGACATTTAGCTGTTCTATGGGAAGCTG 364
||||| : : : : : ||||| : : : : :
Db 129 ArgGluValAspIleLeuAsnHisValPheAspValGluSerPheValSerArgLeu 148
QY 365 GAGAAGGCC-----CAGGCAAAAGACCAGCAGGAAGAATAATTGGGAAAAAACAAG 418
: : : : : ||||| : : : : :
Db 149 GlnLysSerAlaGluAlaAlaArgValLeuGluHisArgGluArgGlyArgArgSerArg 168
QY 419 -----GACCAGGAGGTCTCACC-----CAGGCACAG 445
: : : : : ||||| : : : : :
Db 169 ArgArgAlaAlaGlyGluGlyLeuLeuThrLeuArgAlaLysProProSerGluAlaGlu 188
QY 446 TACATTGACTCTCCCAAGATCAAGTACAGCTTCAACCTCCTCGGGAAGCTGGCCACC 505
||||| : : : : : ||||| : : : : :
Db 189 TyrThrAspValLeuGlnLysIleLysTyrAlaPheSerLeuLeuAlaArgLeuArgGly 208
QY 506 TGGCTGAAGAGACAAAGTCCCTGAGCTCGTACACATCTCTCAAGTCCCTGAACCTC 565
: : : : : ||||| : : : : :
Db 209 AsnIleAlaAspProSerSerProGluLeuLeuHisPheLeuPheGlyProLeuGlnMet 228
QY 566 ATCTTGGCCAGTGGCCCTGAGGCTGGCTAGCAGCCCAAGTGATCTACCCCTCCTCACC 625
: : : : : ||||| : : : : :
Db 229 IleValAsnThrSerGlyProGluPheAlaSerSerValArgArgProHisLeuThr 248
QY 626 CCTAAAGCTATCACTCTACAGTCCTGTCTAAGCCCACTGAGAGTAACCTTTGGATG 685
||||| : : : : : ||||| : : : : :
Db 249 SerAspAlaValAlaLeuLeuArgAspAsnValThrProArgGluAsnGluLeuThrPthr 268
QY 686 GGGTTGGCCCGCAGCTGACCACTAGCCGGCGGCTGAGCAGCGCATGAGCCCTGCCC 745
: : : : : ||||| : : : : :
Db 269 SerLeuGlyAspSerThrArgProGlyLeuGluLeuSerProGluGluGlyProPro 288
QY 746 TACCAACCCACATCTCAGATGACTGGCAACTTCCA-----GAGCCCTCCAGCCACGA 799
: : : : : ||||| : : : : :
Db 289 TyrArgProGluPhePheSerGlyTyrPgluProProValThrAspProGlnSerArgAla 308
QY 800 CCCTTAGGATACCAGGACCTGTTCCCTTCGCGGGGAAATCATAGTTAGGAGCACC 859
: : : : : ||||| : : : : :
Db 309 -----TrpGluAspProValGluLysGlnLeuGlnHisGluArgArgArgGln 325
QY 860 TCACACTTCTCTGAGGAGAAGACACAAACCAT---GACCCTCAGCCTGGGAGCCCAAC 916
||||| : : : : : ||||| : : : : :
Db 326 GlnSerAlaProGlnValAlaValAsnGlyHisArgLspLeuGluPro----- 341
QY 917 TCCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGAAATGCAAGTCTTG----- 970

Db 342 -----GluSerGluProGlnLeuGluSerGluThrAlaGlyLysTrpValLeuCysAsn 359
||||| : : : : : |||||
QY 971 TACAGATTGAAGTAGAACCCACGGAAGTACTGTGTCTCCAGGAGAGAACTGGAG 1030
: : : : : ||||| : : : : :
Db 360 TyrAspPheGlnAlaArgAsnSerSerGluLeuSerValLysGlnArgAspValLeuGlu 379
QY 1031 GTTCTGACCAACAGCAGCGTGGTGGTGAAGTAATGAGCGGACGAGCGGCTAC 1090
||||| : : : : : ||||| : : : : :
Db 380 ValLeuAspAspSerArgLysTrpTrpLysValargaspProAlaGlyGlnGluGlyTyr 399
QY 1091 ATTCAGCAACATCTCTGGAGCCCTACAGCGGG----- 1126
: : : ||| : : : : : |||||
Db 400 ValProTyrAsnIleLeuThrPro---TyrProGlyProArgLeuHisHisSerGlnSer 418
QY 1127 -----ACCCCTGG----- 1135
|||||
Db 419 ProAlaArgSerLeuAsnSerThrProProProProAlaProAlaProAlaProPro 438
QY 1135 ----- 1135
Db 439 ProAlaLeuAlaArgProArgTrpAspArgProArgTrpAspSerCysAspSerLeuAsn 458
QY 1135 ----- 1135
Db 459 GlyLeuAspProSerGluLysGluLysPheSerGlnMetLeuIleValAsnGluGluLeu 478
QY 1136 -----ACCCAGGCCAGTCA---CCCTCTCGGGTT----- 1162
||||| : : : : : |||||
Db 479 GlnAlaArgLeuAlaGlnGlyArgSerGlyProSerArgAlaValProGlyProArgAla 498
QY 1163 -----CCAATGCTTCGACTTAGCTCGAGGCTCAAGAGGTCACAGACTGGCTGCAGCA 1216
||| ||| ||| ||||| |||||
Db 499 ProGluProGlnLeuSerProGlySerAspAlaSerGluValArgAlaTrpLeuGlnAla 518
QY 1217 GAGAACTTCTCCACTGCCCGTGGAGACTTGGTCCCTGACGGGAGCCAGCTACTT 1276
: : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 519 LysGlyPheSerSerGlyThrValAspAlaLeuGlyValLeuThrGlyAlaGlnLeuPhe 538
QY 1277 CGCATAAGACCTGGGAGCTACAGATGCTATGTCCACAGGAGGCCCAATCTGTCC 1336
: : : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db 539 SerLeuGlnArgGluGluLeuArgAlaValSerProGluGluGlyAlaArgValTrpSer 558
QY 1337 CGGTGGAGGCTGTCAGAGGATGCTG 1363
: : : : : ||| : : : |||||
Db 559 GlnValIleValGlnArgSerLeuLeu 567
RESULT 11
Q9D2M6 PRELIMINARY; PRT; 652 AA.
AC Q9D2M6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE 4632407K17Rik protein.
GN 4632407K17Rik.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=SKIN;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whittaker C., Wilming L.,
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohtsuki S.,
 RA Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
 DR EMBL; AK019490; BAB31756.1; -.
 DR HSSP; Q08509; 1A0J.
 DR MGD; MGI:1914675; 4632407K17Rik.
 DR InterPro; IPR001452; SH3.
 DR Pfam; PF00018; SH3; 1.
 DR ProDom; PD000066; SH3; 1.
 DR SMART; SM00326; SH3; 1.
 DR PROSITE; PS50002; SH3; 1.
 KW SH3 domain.
 SQ SEQUENCE 652 AA; 73007 MW; D43DAA59C2D7C638 CRC64;

Alignment Scores:

Pred. No.: 2,04e-36 Length: 652
 Score: 563.00 Matches: 162
 Percent Similarity: 42.81% Conservative: 79
 Best Local Similarity: 28.77% Mismatches: 194
 Query Match: 18.00% Indels: 128
 DB: 11 Gaps: 15

US-09-762-021A-1 (1-1710) x 09D2M6 (1-652)

QY 2 GCAGCGGACTGAAGACCGCTGCAGAGCGTCTGGAGGAGAGCTGGAGCAAGA---- 58
 DB 80 AlaGluLeuIleArgGluAspIleGlnGlyAlaLeuGlnAsnTyrArgSerGlyArgGly 99
 QY 59 ---CCTCGACTTGGAGCCCTTACGCCAGCGCCAGGACAGATGGAGG-----GGCCCT 106
 DB 100 GluArgArgAlaAlaAlaLeuArgAlaThrGlnGluLeuArgGlyAlaSerPro 119
 QY 107 GCTATGGAAGCGCTGCTATGAGAGCGGACGCTATCTGGAGCGGGATCCCTCCA 166
 DB 120 AlaAlaGluThrPro--ProLeuGlnArg----- 128
 QY 167 GAACGCGCCACGAGGACCCCTAGAGCAGCGCTCCACCATCCCA---AGGCCCTG 223
 DB 129 ---ArgProSerValArgLeuValIleAsnThrValGluProSerAlaValArgGlyArg 147
 QY 224 CCACGCCACACCGTGCCTGAGCAACCAAGTGCCTTTACTCTGCTCTCCAGCGCGTCC 283
 DB 148 ProGlnValGluSerIleProGluThrGluGluAlaArgLysProAspGlnAlaArgThr 167
 QY 284 TCTTCCCGGAGACCA----- 301
 DB 168 ThrSerSerAlaAspProThrSerProAspLeuGlyProArgGlyProGluLeuAlaGly 187
 QY 302 -----GAGAGGAGGAGGAGTCTGAACCATGCTCTAAGGACATGAGCTGTTC 352
 DB 188 LeuGlnAlaGluArgAspValAspIleLeuAsnHisValPheAspValGluSerPhe 207
 QY 353 ATGGGAAGCTGGAGAGGCCAGGCAAGACC-----AGCAGAG 394
 DB 208 ValSerArgLeuGlnLysSerAlaGluAlaThrArgValLeuGluHisArgGluArgGly 227
 QY 395 AGAATTTGGGAAAAAACAGAGGACCGAGGAGGTCTC----- 433
 DB 228 ArgArgThrArgArgAlaAlaGlyGluGlyLeuLeuThrLeuArgAlaLysProPro 247
 QY 434 ACCGAGCAGATGACTCTTCCAGAGATCAAGTACAGCTTCAACCTCTCGGGA 493
 DB 248 ThrGluAlaGluThrAspValLeuGlnLysIleLysTyrAlaPheSerLeuAla 267

QY 494 AGGTGGCCACCTGGCTGAAGGAGACAAAGTGCCTGAGCTCGTACACATCTCTTCAG 553
 DB 268 ArgLeuArgGlyAsnIleAlaAsnProSerSerProGluLeuLeuHisPheLeuGly 287
 QY 554 TCCCTGAACCTTCTGCTGCCAGGTGCCCTGAGGTGGCTAGCAGCCCAAGTGATCTCA 613
 DB 288 ProLeuGlnMetIleValAsnThrSerGlyGlyProGluPheAlaLysSerValArgArg 307
 QY 614 CCCCTCTCACCCCTAAAGCTATCAACCTGCTACAGCTCTGTCTAAGCCCACTGAGAT 673
 DB 308 ProHisLeuThrLeuGluAlaValThrLeuLeuArgAspAsnValThrProGlyGluAsn 327
 QY 674 AACCTTTGGATGGGTGGGCCACGCTGACCTAGCCGGCCGACCTAGCAGCGAT 733
 DB 328 GluLeuThrSerLeuGlyAspSerTyrThrCysProGlyValGluLeuProGlu 347
 QY 734 GAGCCCTGCTCTACCAACCCACATTCAGATGACTGCGCACTTCCAGAGCCCTCCAGC 793
 DB 348 GluGlySerProTyrSerProGluPheTyrAsnGlyTrp-----GluProProAla 364
 QY 794 CAAGCACCTTAGGA-----TACCAGGACCTGTGTTCCCTTCGGGGGGAAGTCATAGG 847
 DB 365 ThrAspProGlnGlyArgProTyrGluAspProValGluGlnLeuHisGluLys 384
 QY 848 TTAGGAGCACCTCACACTTCTCAGGAGAAAGACACACACCAACCTGACCTCAGCCTGG 907
 DB 385 ArgArgArgGlnGlnSerAlaProGlnValAlaValAsnGlyGlnGln----- 400
 QY 908 GACCCCACTCCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGAAATGCAAGTC 967
 DB 401 AspProGluLeuGluThrGluSerGlnLeuGluLysAlaArgLysTyrValLeuGly 420
 QY 968 TTGTAGGACTTTGAGCTAGGAACCCAGCGAAGTGCCTGCTCCAGGAGAGAGCTG 1027
 DB 421 AsnTyrAspPheGlnAlaArgAsnGlySerGluLeuSerValLysHisArgAspValLeu 440
 QY 1028 GAGGTTCGAGCACAGCAAGCGGTGGTGGTGAAGAATGAGCGGAGGAGCGGC 1087
 DB 441 GluValLeuAspArgArgLysTyrTrpLysValArgAspHisGlnGlyGlnGly 460
 QY 1088 TACATTCGAAGCAACATCTCTGAGGCC----- 1114
 DB 461 TyrValProTyrAsnIleLeuThrProHisProGlyProGlnValHisArgSerGlnSer 480
 QY 1115 -----CTACAGCGGGGCCCTGG----- 1135
 DB 481 ProAlaArgHisLeuGluThrSerThrProProProProAlaProAlaPro 500
 QY 1135 ----- 1135
 DB 501 ThrGlnValArgProGlnTrpAspSerCysAspSerLeuAsnSerLeuAspProSerGlu 520
 QY 1136 -----ACCCAG 1141
 DB 521 LysGluLysPheSerGlnMetLeuCysValAsnGluLeuGlnSerArgLeuAlaGln 540
 QY 1142 GGCAGCTCA---CCCTCTCGGTT-----CCAATGCTCGCA 1174
 DB 541 GlyArgSerGlyProSerArgValThrProGlyProArgAlaGlnGluProGlnLeuSer 560
 QY 1175 CTTAGCTCAGCGCTGAAGAGTCTACAGACTGGCTGCGAGCAGAGAACTTCTCCACTGCC 1234
 DB 561 ProArgSerGluAlaSerValValArgAlaTrpLeuGlnThrLysGlyPheSerSerGly 580
 QY 1235 ACGGTGAGCACACTTGGTCCCTGACGGGAGCCAGCTACTTCGCATAGACCTGGGGAG 1294
 DB 581 ThrValGluAlaLeuGlyValLeuThrGlyAlaGlnLeuPheSerLeuGlnLysGluGlu 600
 QY 1295 CTACAGATCTGTCTCCAGGAGGCCCAACCTCTCTCCGCGCTGGAGGCTGTCAG 1354
 DB 601 LeuArgAlaValCysProGluGluGlyAlaArgValTyrSerGlnValThrValGlnArg 620
 QY 1355 AGGATGCTG 1363

```

Db      621 AlaLeuLeu 623
:::||||
RESULT 12
Q8R5F8
ID      Q8R5F8      PRELIMINARY;      PRT;      716 AA.
AC      Q8R5F8;
DT      01-JUN-2002 (TREMBLrel. 21, Created)
DT      01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Epidermal growth factor receptor pathway substrate 8 related protein
DE      1.
OS      EPS8r1.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxId=10090;
[1]
RN      SEQUENCE FROM N.A.
RP      SFRAIN=C57BL/6;
RA      Scita G., Confalonieri S., Offenhauser N., Borgonovo A., Tocchetti A.,
RT      Romano P., Di Fiore P.P.;
RT      "Cloning and characterization of novel members of the Eps8 protein
RL      family.";
RL      Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AY074931; AAL76120.1; -.
KW      Receptor.
SQ      SEQUENCE      716 AA;      80028 MW;      E996E664722E885D CRC64;

```

Alignment Scores:	2.07e-36	Length:	716
Pred. No.:	563.00	Matches:	162
Score:		Mismatches:	79
Percent Similarity:	42.81%	Conservative:	194
Best Local Similarity:	28.77%	Indels:	128
Query Match:	18.00%	Gaps:	15
DB:	11		
US-09-762-021A-1 (1-1710) x Q8R5F8 (1-716)			
QY	2	GCAGAGCGACTGAAGACCAGCGCTGCAGAGGCTCTGGAGGAAGAGCTGGAGCAAGA---	58
DB	144	AlaGluLeuIleArgGluAspIleGlnGlyAlaLeuGlnAsnTyrArgSerGlyArgGly	163
QY	59	---CCTGAGCTTGGAGCCCTTCAGCCAGGCCACGACAGATGGAGG-----GGGCTT	106
DB	164	GluArgArgAlaAlaAlaLeuArgAlaThrGlnGluGluLeuArgArgGlyAlaSerPro	183
QY	107	GCTATGGAAGGCCGCTCCCTATATGAGCAGGCCAGCGTATCTGGAGCCGGGATCCCTCCA	166
DB	184	AlaAlaGluThrPro---ProLeuGlnArg-----	192
QY	167	GAACAGCCCCACCAGAGAGCCCTFAGAGCAGCGCTCCACCATCCCA--AGGCCCTGT	223
DB	193	---ArgProSerValArgLeuValIleAsnThrValGluProSerAlaValArgGlyArg	211
QY	224	CCAGCCACCACCACTAGTCCGCCAGAACCAAGTGCTTTACTGCTGCTCCCAAGCGGTCC	283
DB	212	ProGlnValGluSerIleProGluThrGluGluAlaArgLysProAspGlnAlaArgThr	231
QY	284	TCCTCCCCCAGGACCA-----	301
DB	232	ThrSerSerAlaAspProThrSerProAspLeuGlyProArgGlyProGluLeuAlaGly	251
QY	302	-----CAGAGGACGAGGAAGTCTCGAACCATGCTCCTAAGGCATTCAGCTGTC	352
DB	252	LeuGlnAlaGluArgaspValAspIleLeuAsnHisValPheAspValGluSerPhe	271
QY	353	ATGGGAAAGCTGAGAGAGGCCAGGCCAAAGACC-----AGCAGGAAG	394
DB	272	ValSerArgLeuGlnLysSerAlaGluAlaThrArgValLeuGluHisArgGluArgGly	291
QY	395	AAGAAATTTGGGAAAAAACAAGGACCAGGAGGTCTC-----	433


```
Db 577 IleGluGlyArgGlyGlyAlaAlaGlyHisGlyHisGlyHisGlyHisGlyProGly 596
QY 872 -----CAGGAGAACACACACAAACCATGACCT 898
Db 597 AlaGlyProThrLeuSerAlaIleThrAlaGlyLeuGlnAsnLeuHisThrArgGluSer 616
QY 899 CAGCTGGGGACCCCAACTCCAGGCCCTCCAGCCGCCCAACCT----- 940
Db 617 ArgSerGlyAsnGlyTyrGlyAlaGlyAlaGlyProGlyProSerSerGluLeuGlyGly 636
QY 940 ----- 940
Db 637 GlyGlyArgGlyLeuProAsnValSerAspAspGlnMetLeuGluSerTrpLeuGluAsp 656
QY 941 -----GCCAGCCAGCCGTAAGTCTGTAGAGTTTGAAGCTAGGAAACCCA 994
Db 657 LeuGlnAlaThrGlyAlaLysIleValLeuValThrTyrProArgThrAlaAsnAsp 676
QY 995 CGGGAAGTACTGTGTCCAGGAGAGAGCTGGAGTTCTGGACACAGCAAGCGGTGG 1054
Db 677 LysGluLeuSerValMetArgGlyGluTyrLeuGluIleLeuAspThrArgLysTrp 696
QY 1055 TGGCTGGTGAAGATCAGCGGGAGCGGAGCGCTACATTCAGCAACATCTGGAGCCC 1114
Db 697 TrpLysAlaArgAsnMetArgGlyGlnValAlaHisValProHisThrIleValThrPro 716
QY 1115 CTA-----CAGCCGGGGACC 1129
Db 717 PheAsnPheGlyAspGlyAlaGlnPheTyrGlyGlnGlnGlnProThr 736
QY 1130 -----CCTGGGACCCAGGGCCAG 1147
Db 737 GlyProThrGlyProGlyAsnLysSerArgSerGlyAspAsnProGlyMetGluGlnArg 756
QY 1148 TCACCTCTCGGGTTCCTCAATGCTTCCA 1174
Db 757 SerProAspProThrAspMetMetArg 765
RESULT 14
Q90329 PRELIMINARY: PRT; 778 AA.
AC Q90329;
DT 01-MAY-2000 (TRENBLrel. 13, Created)
DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TRENBLrel. 20, Last annotation update)
DE AROUSER.
GN ARU OR CG4276.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidae; Drosophilidae; Drosophila.
OX NCBI_taxid=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA Liu C.H., Bourne J., Harris S.J., O'Kane C.J., Moffat K.G.;
RT "Drosophila epidermal growth factor receptor kinase substrate Eps8.";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AF208262; AAF21013.1; -.
DR HSSP; Q08509; IAOJ.
DR FlyBase; FBgn0029095; aru.
DR InterPro; IPR000050; PID_domain.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00018; SH3; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00462; PTB; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
KW SH3 domain. 778 AA; 85132 MW; E0ED1485F5285189 CRC64;
SQ SEQUENCE
```

Alignment Scores:

```
Pred. No.: 1.42e-23 Length: 778
Score: 402.00 Matches: 122
Percent Similarity: 37.31% Conservative: 53
Best Local Similarity: 26.01% Mismatches: 122
Query Match: 12.86% Indels: 172
DB: 5 Gaps: 13
US-09-762-021A-1 (1-1710) x Q90329 (1-778)
QY 281 TCCTCTTCCCGAGGAGCCAGAGGAGGAGTCTCAACCATGCTCCTAAGGAC 340
Db 298 SerThrSerSerGluLysTyrGluArgAspValAlaValLeuAsnHisCysPheAsp 317
QY 341 ATTGAGCTGTTTCATGGAAAGCTGGAGAGGCCCGAGCA-----AAG 382
Db 318 IleGluLysPheIleAlaArgLeuGlnHisAlaAlaAlaSerArgGluLeuGluArg 337
QY 383 ACCACGACGAGAAAGAAATTTGGGAAAAAACAAGGACCGAGGAGGTCTACC----- 436
Db 338 ArgArgArgAsnArgLysSerLysLysArgAspProGlyGluGlyLeuLeuThrLeuArg 357
QY 437 -----CAGGCACAGTACATTGACTGCTTCCACAAGATCAAGTACAGCTTC 481
Db 358 ThrArgProHisGluLysGluPheValAspIlePheAlaLysPheLysLeuSerPhe 377
QY 482 AACCTCTCTGGGAGGCTGGCCACCTGGCTGGAAGGAGACAAGTCCCTCAGCTCGTACAC 541
Db 378 AsnLeuLeuAlaLysLeuLysAlaHisIleHisAspProAsnAlaProGluLeuValHis 397
QY 542 AFCCTCTTCAAGTCCCTGAACCTTCATCTCCGCGGAGGAGTGCCT-----CAGGCT 589
Db 398 PheLeuPheThrProLeuAlaLeuIleValGluAlaSerSerAspThrTyrTyrGluSer 417
QY 590 GGCCTAGGAGCCCAAGTATCTACCCCTCTCACCCTTCAACCTTAAAGCTATCAACCTGCTACAG 649
Db 418 GlnLeuProAlaArgValValAsnProLeuLeuThrArgGluAlaIleAsnLeuLeuIle 437
QY 650 TCCTGCTTAAGCCACCTGAGAGTAACTTGGTGGGTGGGCCCGACCTGGACCACT 709
Db 438 AsnCysValThrSerLysGluThrGluLeuThrArgSerLeuGlyAspAlaTrpValIle 457
QY 710 AGCCGGCGGACTGGACGAGCGATGAGCCCTGCCCTACCAACCCACATTCATCATGATGAC 769
Db 458 ProArgAspGlnTrpLys---AspAspValGlySerTyrHisProValPheLeuAspGly 476
QY 770 TGG-----CAACTTCAGAGCCCTCCAGCCACCAACCCCTTA 805
Db 477 TrpSerProAspTyrIleAsnAspGluLeuGluProProAsnSerProAla 496
QY 805 ----- 805
Db 497 HisValSerLysArgArgLeuGluValGlnAlaGlyProGlyLeuAsnGlyArgGlyGly 516
QY 806 ---GGATACCGAGGACCTGTTTC----- 826
Db 517 GlyGlyTyrAspAspTyrAspSerGlyAsnGlyMetAsnMetAlaMetGlyIleGluLys 536
QY 827 -----CTTCGG-----CGGGAGAGTCA 844
Db 537 TyrThrIleHisHisGlyAsnAspGlyLeuArgGluArgGluArgGluArgGluArgGlu 556
QY 845 AGGTTTAGGAGCACCTCA----- 862
Db 557 ArgAlaGlyAlaIleSerAlaSerAspPheAsnAlaArgSerGluLeuSerPheAspSer 576
QY 863 -----CATTCTCTCT----- 871
Db 577 IleGluGlyArgGlyGlyAlaAlaGlyHisGlyHisGlyHisGlyHisGlyProGly 596
QY 872 -----CAGGAGACACACACATCAACCTCC 898
Db 597 AlaGlyProThrLeuSerAlaIleThrAlaGlyLeuGlnAsnLeuHisThrArgGluSer 616
```



```

Db 376 ArgCysSerSerThrArgAsnCysArgArg-AlaTirProArgAlaAlaArgAspAr 395
QY 1385 GCTTAGACACCTCCAAGAACCCAGCCCC 1412
Db 395 gAlaAlaGlnSerGlnGlyProAlaPro 404
```

Search completed: February 25, 2003, 19:43:01
Job time : 73 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 25, 2003, 07:32:52 ; Search time 1688 Seconds
(without alignments)
16406.565 Million cell updates/sec

Title: US-09-762-021A-1
Perfect score: 1710
Sequence: 1 ggcagcgactgaagacca.....aaaaaaaaaataangataaa 1710

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0
Searched: 16154066 seqs, 8097743376 residues
Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

- 1: em_estba:**
- 2: em_esthum:**
- 3: em_estin:**
- 4: em_estmu:**
- 5: em_estov:**
- 6: em_estpl:**
- 7: em_estro:**
- 8: em_htc:**
- 9: gb_est1:**
- 10: gb_est2:**
- 11: gb_htc:**
- 12: gb_est3:**
- 13: gb_est4:**
- 14: gb_est5:**
- 15: em_estfun:**
- 16: em_estom:**
- 17: gb_gss:**
- 18: em_gss_hum:**
- 19: em_gss_inv:**
- 20: em_gss_pln:**
- 21: em_gss_vrt:**
- 22: em_gss_fun:**
- 23: em_gss_mam:**
- 24: em_gss_mus:**
- 25: em_gss_other:**
- 26: em_gss_pro:**
- 27: em_gss_rod:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	662	38.7	662	14	BM824107
2	600.8	35.1	604	14	BM744431
3	587.4	34.4	589	14	BM742903
4	560.4	32.8	562	14	BM756251
5	547.4	32.0	549	14	BM770581
6	542.4	31.7	556	10	AW778776

7	530.4	31.0	532	14	BM769901
8	517.6	30.3	604	14	BM741405
c	496.4	29.0	507	12	BF001871
	491.8	28.8	551	10	AW841643
	477.2	27.9	483	14	BM797626
	473.2	27.7	478	14	BM792701
c	472	27.6	534	10	AW937714
c	454.4	26.6	456	9	AI339104
c	443.6	25.9	901	14	BQ951422
	443.6	25.9	901	14	BQ951422
	417	24.4	521	14	BM824199
	413.8	24.2	431	14	BQ320003
c	409.4	23.9	411	9	AA861042
c	404.4	23.6	417	14	BM757009
	395	23.1	427	14	BM818771
c	389	22.7	401	9	AA134985
c	378.6	22.1	728	13	BI284392
c	374.8	21.9	378	9	AA868144
c	366.6	21.4	773	13	BG966755
c	356.6	20.9	685	9	AI746552
c	354.6	20.7	419	14	BQ326654
c	352.2	20.6	422	13	BG958438
c	344.6	20.2	512	9	AA710419
c	341.6	20.0	449	13	BM445779
c	332.2	19.4	356	12	BF755318
c	331.6	19.4	672	10	AV378666
c	325.8	19.1	413	9	AA134946
c	322.4	18.9	329	13	BG982594
c	321.2	18.8	606	17	AQ022003
c	319.2	18.7	369	9	AI626100
c	312.4	18.3	385	14	BQ320515
c	309.4	18.1	322	13	BG982587
c	307.6	18.0	468	12	BF334476
c	305.2	17.8	310	9	AA922724
c	295.2	17.3	524	9	AA535447
c	279.8	16.4	509	9	AI429197
c	274.2	16.0	362	9	AA135023
c	265.6	15.5	368	10	AW951951
c	263.4	15.4	271	10	AW818756
c	263.4	15.4	271	10	AW818837

ALIGNMENTS

RESULT 1	BM824107	BM824107	662 bp	mRNA	linear	EST 06-MAR-2002
LOCUS	BM824107	K-EST0095517	S22SNU16n1	Homo sapiens	cDNA clone	S22SNU16n1-89-A05
DEFINITION	5', mRNA sequence.					
ACCESSION	BM824107					
VERSION	BM824107.1	GI:19180520				
KEYWORDS	EST.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.					
TITLE	21C Frontier Korean EST Project 2001					
JOURNAL	Unpublished (2002)					
COMMENT	Contact: Kim YS Genome Research Center Korea Research Institute of Bioscience & Biotechnology 52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea Tel: +82-42-860-4470 Fax: +82-42-860-4409 Email: yongsung@mail.kribb.re.kr Plate: 89 row: A column: 05 High quality sequence stop: 662. Location/Qualifiers 1. .662					
FEATURES						
source						


```
QY 591 GCCTAGCAGCCCAAGTGATCTCACCCCTCCTCACCCCTAAAGCTATCAACCTGTACAGT 650
|||||
Db 181 GCCTAGCAGCCCAAGTGATCTCACCCCTCCTCACCCCTAAAGCTATCAACCTGTACAGT 240
QY 651 CCTGTCTAAGCCCACTCAGAGTAACCTTTGGATGGGGTTGGGCCAGCCTGGACCACTA 710
|||||
Db 241 CCTGTCTAAGCCCACTCAGAGTAACCTTTGGATGGGGTTGGGCCAGCCTGGACCACTA 300
QY 711 GCCGGGCCAGCTGGACAGCGGATGAGCCCTCCCTACCAACCCACATCTCAGATGACT 770
|||||
Db 301 GCCGGGCCAGCTGGACAGCGGATGAGCCCTCCCTACCAACCCACATCTCAGATGACT 360
QY 771 GCGAACTTCCAGAGCCCTCCAGCCCAAGCACCTTTAGGATACCAAGACCTCTGTTCCTTC 830
|||||
Db 361 GCGAACTTCCAGAGCCCTCCAGCCCAAGCACCTTTAGGATACCAAGACCTCTGTTCCTTC 420
QY 831 GCCGGGGAGTCATAGTTAGGAGCAGCTCACACTTTCCTCAGGAGAGACACACACACC 890
Db 421 GCCGGGGAGTCATAGTTAGGAGCAGCTCACACTTTCCTCAGGAGAGACACACACACC 480
QY 891 ATGACCCCTCAGCCTGGGACCCCAACTCCAGGCCCTCCAGCCCAAACTCCGCCAGCCAG 950
|||||
Db 481 ATGACCCCTCAGCCTGGGACCCCAACTCCAGGCCCTCCAGCCCAAACTCCGCCAGCCAG 540
QY 951 CCCTGANAATCAAGTCTTGTACGAGTTTGAAGCTTAGGAACCCACGGGAAGTGTGTGG 1010
|||||
Db 541 CCCTGANAATCAAGTCTTGTACGAGTTTGAAGCTTAGGAACCCACGGGAAGTGTGTGG 600
QY 1011 TCCA 1014
Db 601 TCCA 604

RESULT 3
LOCUS BM742903
DEFINITION K-EST0015941 S6SNUG20 Homo sapiens cDNA clone S6SNUG20-10-H03 5',
mRNA sequence.
ACCESSION BM742903
VERSION BM742903.1 GI:19064232
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 589)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 10 row: H column: 03
High quality sequence stop: 589.
Location/Qualifiers
1. 589
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S6SNUG20-10-H03"
/clone_lib="S6SNUG20"
/sex="F"
/tissue_type="Ascites"
/cell_type="Scattering floating"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pCNS; Site_1: EcoRI;
```

```
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
contained cDNA vectors were used for transfection of
competent cells E. coli Top10F' by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."

BASE COUNT 144 a 201 c 139 g 105 t
ORIGIN

Query Match 34.4%; Score 587.4; DB 14; Length 589;
Best Local Similarity 99.8%; Pred. No. 1.le-136;
Matches 588; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 411 AAACAAGGACCCAGGAGGTCTCACCCAGGCACAGTACATTGACTGCTTCCAGAGATCA 470
|||||
Db 1 AAACAAGGACCCAGGAGGTCTCACCCAGGCACAGTACATTGACTGCTTCCAGAGATCA 60
QY 471 AGTACAGCTTCAACCTCCTGGGAAGGCTGGCCACTGGCTGAAGGAGACAAGTGGCCCTG 530
|||||
Db 61 AGTACAGCTTCAACCTCCTGGGAAGGCTGGCCACTGGCTGAAGGAGACAAGTGGCCCTG 120
QY 531 AGCTCGTACACATCCTCTTCAAGTCCCTGAACTTATCTATCTGCGCCAGGTGCGCTGAGGCTG 590
|||||
Db 121 AGCTCGTACACATCCTCTTCAAGTCCCTGAACTTATCTATCTGCGCCAGGTGCGCTGAGGCTG 180
QY 591 GCCTAGAGCCCAAGTATCTACCCCTCCTCACCCCTAAAGCTATCAACCTGTACAGT 650
|||||
Db 181 GCCTAGAGCCCAAGTATCTACCCCTCCTCACCCCTAAAGCTATCAACCTGTACAGT 240
QY 651 CTTGTCTAAGCCCACTCAGAGTAACTTTGGATGGGGTTGGGCCAGCCTGGACCACTA 710
|||||
Db 241 CTTGTCTAAGCCCACTCAGAGTAACTTTGGATGGGGTTGGGCCAGCCTGGACCACTA 300
QY 711 GCCGGGCCAGCTGGACAGCGGATGAGCCCTCGCCTACCAACCCACATTTCTCAGATGACT 770
|||||
Db 301 GCCGGGCCAGCTGGACAGCGGATGAGCCCTCGCCTACCAACCCACATTTCTCAGATGACT 360
QY 771 GCGAACTTCCAGAGCCCTCCAGCCCAAGCACCTTTAGGATACCAAGACCTCTGTTCCTTC 830
|||||
Db 361 GCGAACTTCCAGAGCCCTCCAGCCCAAGCACCTTTAGGATACCAAGACCTCTGTTCCTTC 420
QY 831 GCCGGGGAAGTCATAGTTAGGAGCAGCTCACACTTTCCTCAGGAGAGACACACACACC 890
|||||
Db 421 GCCGGGGAAGTCATAGTTAGGAGCAGCTCACACTTTCCTCAGGAGAGACACACACACC 480
QY 891 ATGACCCCTCAGCCTGGGAGCCCCCAACTCCAGGCCCTCCAGCCCAAACTCCGCCAGCCAG 950
|||||
Db 481 ATGACCCCTCAGCCTGGGAGCCCCCAACTCCAGGCCCTCCAGCCCAAACTCCGCCAGCCAG 540
QY 951 CCCTGANAATCAAGTCTTGTACGAGTTTGAAGCTTAGGAACCCACGGGAAGTGTGG 999
|||||
Db 541 CCCTGANAATCAAGTCTTGTACGAGTTTGAAGCTTAGGAACCCACGGGAAGTGTGG 589

RESULT 4
LOCUS BM756251
DEFINITION K-EST0034532 S6SNUG20 Homo sapiens cDNA clone S6SNUG20-28-D01 5',
mRNA sequence.
ACCESSION BM756251
VERSION BM756251.1 GI:19085866
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
```

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 562)
Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R., Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and Kim,Y.S.

21C Frontier Korean EST Project 2001
Unpublished (2002)
Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409

Email: yongsung@mail.kr.ibm.re.kr

Plate: 28 row: D column: 01

High quality sequence stop: 562.

Location/Qualifiers

source

1..562

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="S6SNU620-28-D01"

/clone_lib="S6SNU620"

/sex="F"

/tissue_type="Ascites"

/cell_type="Scattering floating"

/cell_line="SNU-620"

/lab_host="Top10F"

/note="Organ: Stomach; Vector: pcNS; Site_1: EcoRI; Site_2: NotI; The poly (A)+ RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tabacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including EcoR I site by treatment of T4 RNA ligase and the first strand cDNA was synthesized from oligo dt-selected mRNA by priming with dt-tailed vector. The dt-tailed vector was adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

136 a 195 c 130 g 101 t

Query Match 32.8%; Score 560.4; DB 14; Length 562;

Best Local Similarity 99.8%; Pred. No. 6.8e-130;

Matches 561; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 411 AAAACAAGACGAGGAGGCTCACCCAGGCACAGTACATTGCTTCCAGAGATCA 470

Db 1 AAACAAGACGAGGAGGCTCACCCAGGCACAGTACATTGCTTCCAGAGATCA 60

QY 471 AGTACAGCTTCAACCTCTCGGAAGCGTGGCCACCTGGCTGAAGGAGACAAAGTGGCCCTG 530

Db 61 AGTACAGCTTCAACCTCTCGGAAGCGTGGCCACCTGGCTGAAGGAGACAAAGTGGCCCTG 120

QY 531 AGTGTGTACATCTCTTCAAGTTCCTGAATTCATCTCGGCGAGGTCCTTGAGGCTG 590

Db 121 AGTGTGTACATCTCTTCAAGTTCCTGAATTCATCTCGGCGAGGTCCTTGAGGCTG 180

QY 591 GCTAGACGCCCAAGTGATCTCACCCCTCTCACCCCTTAAGCTATCAACCTGCTACAGT 650

Db 181 GCTAGACGCCCAAGTGATCTCACCCCTCTCACCCCTTAAGCTATCAACCTGCTACAGT 240

QY 651 CTGTGTAAAGCCCACTGAGAGTAACCTTTGGATGGGGTTGGCCCGAGCCTGGACCACTA 710

Db 241 CTGTGTAAAGCCCACTGAGAGTAACCTTTGGATGGGGTTGGCCCGAGCCTGGACCACTA 300

QY 711 GCGGGGCCACTGGACGCGATGAGCCCTCGCCCTACCAACCCACATTCAGATGACT 770

|||||

Db 301 GCCGGCGGACTGGACAGCGGATGAGCCCTGCCCTACCAACCCACATTTCTCGGATGACT 360
QY 771 GCGCACTTCCAGAGCCCTCCACGCAAGCACCCCTTAGGATACCAGGACCCCTGTTCCCTTC 830
Db 361 GCGCACTTCCAGAGCCCTCCACGCAAGCACCCCTTAGGATACCAGGACCCCTGTTCCCTTC 420
QY 831 GCGGGGAAGTCATAGGTTAGGGAGCACCTCACACTTTCTCAGGAGAGACACACAACC 890
Db 421 GCGGGGAAGTCATAGGTTAGGGAGCACCTCACACTTTCTCAGGAGAGACACACAACC 480
QY 891 ATGACCCCTCAGCTGGGGACCCCACTCCAGGCCCTCCAGCCCCCAACCTGCCAGCCAG 950
Db 481 ATGACCCCTCAGCTGGGGACCCCACTCCAGGCCCTCCAGCCCCCAACCTGCCAGCCAG 540
QY 951 CCCTGAAATGCAAGTCTTGTA 972
Db 541 CCCTGAAATGCAAGTCTTGTA 562

RESULT 5

BM770581

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

21C Frontier Korean EST Project 2001

Unpublished (2002)

Contact: Kim YS

Genome Research Center

Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea

Tel: +82-42-860-4470

Fax: +82-42-860-4409

Email: yongsung@mail.kr.ibm.re.kr

Plate: 16 row: H column: 08

High quality sequence stop: 549.

Location/Qualifiers

1..549

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="S6SNU620s1-16-H08"

/clone_lib="S6SNU620s1"

/sex="F"

/tissue_type="Ascites"

/cell_type="Scattering floating"

/cell_line="SNU-620"

/lab_host="Top10F"

/note="Organ: Stomach; Vector: pcNS; Site_1: EcoRI;

Site_2: NotI; The poly (A)+ RNA was dephosphorylated with

bacterial alkaline phosphatase (BAP) and then decapped

with tabacco acid pyrophosphatase (TAP). The decapped

intact mRNA was ligated with DNA-RNA linker including EcoR

I site by treatment of T4 RNA ligase and the first strand

cDNA was synthesized from oligo dt-selected mRNA by

priming with dt-tailed vector. The dt-tailed vector was

adjusted to have about 60nt. The cDNA vector was

circularized with E. coli DNA ligase after digestion of

EcoRI which site is also included in vector. An RNA strand

converted to a DNA strand by Okayama-Berg method. The

obtained cDNA vectors were used for transformation of

competent cells E. coli Top10F by electroporation method.

The cDNA libraries constructed by this method are

full-length enriched cDNA library. After analyzing and

sequencing about 2,000 ~ 3,000 colonies in original cDNA library, the abundant cDNAs were selected and amplified by PCR reaction using vector region primer including T7 promoter as 5' primer and N(dT)14 as 3' primer. The PCR products were used as template for synthesis of biotinylated single stranded RNA by in vitro transcription reaction. The synthesized RNA probes were hybridized with antisense single stranded cDNAs prepared from original library and incubated with avidin-gel. After removing DNA-RNA hybrids by centrifuge, the subtracted cDNA libraries were constructed by transfection of the remaining DNA into competent cells *E. coli* Top10⁺ with electroporation method.

```
BASE COUNT      133 a   193 c   127 g   96 t
ORIGIN

Query Match      32.0%; Score 547.4; DB 14; Length 549;
Best Local Similarity 99.8%; Pred. No. 1.3e-126;
Matches 548; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 411 AAAACAAGCAGCAGGAGGTCTCACCCAGGCACAGTACATTCAGTCTCCAGAGATCA 470
      |||
Db 1 AAAACAAGCAGCAGGAGGTCTCACCCAGGCACAGTACATTCAGTCTCCAGAGATCA 60

Qy 471 AGTACAGCTTCAACCTCTCGGAAGCTGGCCACCTGCTGAAGGAGACAAGTGCCCTG 530
      |||
Db 61 AGTACAGCTTCAACCTCTCGGAAGCTGGCCACCTGCTGAAGGAGACAAGTGCCCTG 120

Qy 531 AGCTGTACACATCTCTTCAAGTCCCTGAATTCATCTCTGGCCAGGTGCCCTGAGGCTG 590
      |||
Db 121 AGCTGTACACATCTCTTCAAGTCCCTGAATTCATCTCTGGCCAGGTGCCCTGAGGCTG 180

Qy 591 GCCTAGCAGCCCAAGTGTATCTCACCCCTCCTCACCCCTAAAGCTATCAACCTGTACAGT 650
      |||
Db 181 GCCTAGCAGCCCAAGTGTATCTCACCCCTCCTCACCCCTAAAGCTATCAACCTGTACAGT 240

Qy 651 CCTGTCTAAGCCCACTGAGAGTAACTTTTGGATGGGTTGGGCCAGCCTGGACCACTA 710
      |||
Db 241 CCTGTCTAAGCCCACTGAGAGTAACTTTTGGATGGGTTGGGCCAGCCTGGACCACTA 300

Qy 711 GCCGGCCGACTGGACAGGCGATGAGCCCTGCTTACCAACCCACATCTTCAGATGACT 770
      |||
Db 301 GCCGGCCGACTGGACAGGCGATGAGCCCTGCTTACCAACCCACATCTTCAGATGACT 360

Qy 771 GGCAACTTCCAGAGCCCTCCAGCAGCAGCCCTTAGGATACAGGACCCCTGTTCCCTTC 830
      |||
Db 361 GGCAACTTCCAGAGCCCTCCAGCAGCAGCCCTTAGGATACAGGACCCCTGTTCCCTTC 420

Qy 831 GCGGGGGAAGTCATAGTTAGGAGCAGCTTCCCTCAGGAGAGACACACACAC 890
      |||
Db 421 GCGGGGGAAGTCATAGTTAGGAGCAGCTTCCCTCAGGAGAGACACACACAC 480

Qy 891 ATGACCTTCAGCCTGGGAGCCCAACTCCAGGCGCTCAGGCCCCAAACCTGCCAGCCAG 950
      |||
Db 481 ATGACCTTCAGCCTGGGAGCCCAACTCCAGGCGCTCAGGCCCCAAACCTGCCAGCCAG 540

Qy 951 CCCTGAAAA 959
      |||
Db 541 CCCTGAAAA 549
```

```
RESULT 6
AW78776/c
LOCUS      AW78776
DEFINITION hol2h03.x1 NCI_CGAP_Col4 Homo sapiens cDNA clone IMAGE:3037205 3'
            similar to SM:EP58_MOUSE Q08509 EPIDERMAL GROWTH FACTOR RECEPTOR
            KINASE SUBSTRATE EPS8. ;, mRNA sequence.
ACCESSION  AW78776
VERSION    AW78776.1 GI:7793366
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 556)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
Tissue-CGAP National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue-Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone Distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
infoimage.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 413.
Location/Qualifiers
1..556
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3037205"
/clone_lib="NCI_CGAP_Col4"
/tissue_type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/note="Organ: colon; Vector: pCMV-Sport6; Site:1: Salt;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.7 kb. Life Technologies catalog #:
11531-019"

```
BASE COUNT      117 a   131 c   168 g   140 t
ORIGIN

Query Match      31.7%; Score 542.4; DB 10; Length 556;
Best Local Similarity 99.8%; Pred. No. 2.3e-125;
Matches 543; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1151 CCCTCTCGGGTTCCTCAATGCTTCGACTTAGCTCGAGGCTGAAGAGGTCCAGACTGGCTG 1210
      |||
Db 544 CCCTCTCGGGTTCCTCAATGCTTCGACTTAGCTCGAGGCTGAAGAGGTCCAGACTGGCTG 485

Qy 1211 CAGGAGAGAACTTCTCCACTGCCACGGTGAGGACACTTGGGTCCCTGACGGGGAGCCAG 1270
      |||
Db 484 CAGGAGAGAACTTCTCCACTGCCACGGTGAGGACACTTGGGTCCCTGACGGGGAGCCAG 425

Qy 1271 CTACTTCGCATTAAGACCTTGGGGAGCTACAGATGCTATGTCCACAGGAGGCCCAAGATC 1330
      |||
Db 424 CTACTTCGCATTAAGACCTTGGGGAGCTACAGATGCTATGTCCACAGGAGGCCCAAGATC 365

Qy 1331 CTGTCCCGGCTGGAGGCTGTCAGAAAGATGCTGGGGATAAGCCCTTAGGCACCACTTTAG 1390
      |||
Db 364 CTGTCCCGGCTGGAGGCTGTCAGAAAGATGCTGGGGATAAGCCCTTAGGCACCACTTTAG 305

Qy 1391 ACACCTCCAAAGAACAGGCCCGCTGTATGCAAGATGGCAGATCTGTATACCCATTAGAGCC 1450
      |||
Db 304 ACACCTCCAAAGAACAGGCCCGCTGTATGCAAGATGGCAGATCTGTATACCCATTAGAGCC 245

Qy 1451 CCGAGAATTCCTCTTCTGGATCCCGAGTTTGGCAGCAAAACCCACACCCAGCTCACACAGC 1510
      |||
Db 244 CCGAGAATTCCTCTTCTGGATCCCGAGTTTGGCAGCAAAACCCACACCCAGCTCACACAGC 185

Qy 1511 AAAACAATGGACAGGCCAGAGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGAG 1570
      |||
Db 184 AAAACAATGGACAGGCCAGAGGCTGAAGCAAAACAGTGTCCCTTCTGGCTGTGTGGAG 125

Qy 1571 CCTCCCCAGTAACCACTTATTTATTTACCTCTTTTCCCAAAACCTGGAGCATTTATGCCTA 1630
      |||
Db 124 CCTCCCCAGTAACCACTTATTTATTTACCTCTTTTCCCAAAACCTGGAGCATTTATGCCTA 65

Qy 1631 GCCTTGTCAAGAACTGTTCAGTCCCTCTCTTCTCAATAAAAGCATCTTCAAGCTTGTGA 1690
      |||
Db 64 GCCTTGTCAAGAACTGTTCAGTCCCTCTCTTCTCAATAAAAGCATCTTCAAGCTTGTGA 5
```

```
QY 1691 AAAA 1694
|||||
Db 4 AAAA 1

RESULT 7
BM769901
LOCUS K-EST0053246 S14K402 Homo sapiens cDNA clone S14K402-24-B06 5', linear EST 04-MAR-2002
DEFINITION mRNA sequence.
ACCESSION BM769901
VERSION BM769901.1 GI:19099516
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE 1 (bases 1 to 532)
AUTHORS Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
Kim,Y.S.
TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 24 row: B column: 06
High quality sequence stop: 532.
Location/Qualifiers
1..532
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="S14K402-24-B06"
/clone_lib="S14K402"
/cell_line="K402"
/lab_host="Top10F"
/note="Organ: Stomach; Vector: pTZ18Rp1; Site_1: EcoRI;
Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
bacterial alkaline phosphatase (BAP) and then decapped
with tabacco acid pyrophosphatase (TAP). The decapped
intact mRNA was ligated with DNA-RNA linker including EcoR
I site by treatment of T4 RNA ligase and the first strand
cDNA was synthesized from oligo dt-selected mRNA by
priming with dt-tailed vector. The dt-tailed vector was
adjusted to have about 60nt. The cDNA vector was
circularized with E. coli DNA ligase after digestion of
EcoRI which site is also included in vector. An RNA strand
converted to a DNA strand by Okayama-Berg method. The
obtained cDNA vectors were used for transfection of
competent cells E. coli Top10F by electroporation method.
The cDNA libraries constructed by this method are
full-length enriched cDNA library."
BASE COUNT 123 a 187 c 125 g 97 t
ORIGIN
Query Match 31.0%; Score 530.4; DB 14; Length 532;
Best Local Similarity 99.8%; Pred. No. 2.3e-122;
Matches 531; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 475 CAGCTTCAACCTCCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCTGAGCT 534
|||||
Db 1 CAGCTTCAACCTCCTGGGAAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCTGAGCT 60
|||||

QY 535 CGTACATCCTCTTCAAGTCCCTGAACCTCATCTCTGGCCAGGTGCGCTGAGGCTGGCCT 594
|||||
Db 61 CGTACATCCTCTTCAAGTCCCTGAACCTCATCTCTGGCCAGGTGCGCTGAGGCTGGCCT 120
|||||

QY 595 AGCAGCCCAAGTGATCTACCCCTCTCTACCCCTAAAGCTATCAACCTGCTACAGTCTCG 654
|||||
```

adjusted to have about 60nt. The cDNA vector was circularized with E. coli DNA ligase after digestion of EcoRI which site is also included in vector. An RNA strand converted to a DNA strand by Okayama-Berg method. The obtained cDNA vectors were used for transfection of competent cells E. coli Top10F' by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

BASE COUNT	147 a	207 c	142 g	108 t
ORIGIN				
Query Match	30.38;	Score 517.6;	DB 14;	Length 604;
Best Local Similarity	97.58;	Pred. No. 3.9e-119;		
Matches 589;	Conservative 0;	Mismatches 9;	Indels 6;	Gaps 6;
Qy	410	AAAAACAAGACCCAGGAGGTCTACCCAGGCACACAGTACATTTGACTGCTTCCAGAAATC	469	
Db	1	AAAAACAAGACCCAGGAGGTCTACCCAGGCACACAGTACATTTGACTGCTTCCAGAAATC	60	
Qy	470	AAGTACAGCTTCAACCTCTCTGGAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCCT	529	
Db	61	GAGTACAGCTTCAACCTCTCTGGAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCCT	120	
Qy	530	GAG-CTCGTACACATCTT-CTTCAAGTCCCT-GAAGTTCATCTT-GGCGAGGTGCCCTG-	584	
Db	121	GAGACTCTGACACATCTTCTCAAGTCCCTAGAATCTTCACTCTGGGCGAGGTGCCCTG	180	
Qy	585	AGGCTGGCTAGCAGCCCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGC	644	
Db	181	AGGCTGGCTAGCAGCCCAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGC	240	
Qy	645	TACAGTCTGTCTAAGCCCACTGAGAGTAACCTTTGGATGGGTTGGGCCAGCCTGGA	704	
Db	241	TACAGTCTGTCTAAGCCCACTGAGAGTAACCTTTGGATGGGTTGGGCCAGCCTGGA	300	
Qy	705	CACTAG-CCGGCGGACTGGACAGCGGATGAGCCCTTACCCCTTACCAACCCACATTC	763	
Db	301	CACTAGTCCGGCGGACTGGACAGCGGATGAGCCCTTACCCCTTACCAACCCACATTC	360	
Qy	764	GATGACTGGCACTTCCAGAGCCCTCCAGCAAGCACCTTTAGGATACCCAGGACCTGTT	823	
Db	361	GATGACTGGCACTTCCAGAGCCCTCCAGCAAGCACCTTTAGGATACCCAGGACCTGTT	420	
Qy	824	TCCCTTCGGCGGGAGTCATAGGTTAGGAGCACCTTCACACTTTCTTCAGGAGAAGACA	883	
Db	421	TCCCTTCGGCGGGAGTCATAGGTTAGGAGCACCTTCACACTTTCTTCAGGAGAAGACA	480	
Qy	884	CACACCATGACCTTCAGCTGGGAGCCCACTCCAGGCGCTTCAGGCCCCAAACCTGCC	943	
Db	481	CACACCATGACCTTCAGCTGGGAGCCCACTCCAGGCGCTTCAGGCCCCAAACCTGCC	540	
Qy	944	CAGCCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCTAGGAACCCACGGGA	1003	
Db	541	CAGCCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCTAGGAACCCACGGGA	600	
Qy	1004	ACTG 1007		
Db	601	ACTG 604		

RESULT 9
BF001871/c
LOCUS 7994f11.x1 NCI-CCAP Col6 Homo sapiens cDNA clone IMAGE:3314157 3'
DEFINITION similar to SW:EP8_MOUSE Q08509 EPIDERMAL GROWTH FACTOR RECEPTOR
KINASE SUBSTRATE EP8F. ; mRNA sequence.
ACCESSION BF001871
VERSION BF001871.1 GI:10702146
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 507)
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
AUTHORS National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CCAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL, send email to:
info@image.llnl.gov
Seq primer: -40UP from Gibco
High quality sequence stop: 468.
Location/Qualifiers
1. 507
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3314157"
/clone_lib="NCI-CCAP_Col6"
/tissue_type="colon tumor, RER+"
/lab_host="DH10B"
/note="Organ: colon; Vector: pT73D-Pac (Pharmacia) with a
modified polylinker; Site.1: Not I; Site.2: Eco RI;
Plasmid DNA from the normalized library NCI-CCAP_Col6 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 1057416-1061255, and 1144584-1145351).
Subtraction by Bento Soares and M. Fatima Bonaldo. "
BASE COUNT 108 a 121 c 149 g 128 t 1 others
ORIGIN

Query Match	29.08;	Score 496.4;	DB 12;	Length 507;
Best Local Similarity	98.68;	Pred. No. 8e-114;		
Matches 500;	Conservative 0;	Mismatches 7;	Indels 0;	Gaps 0;
Qy	1181	TCGAGGCTGAAGAGTTCACAGACTGGCTGCAGGAGAGAACTTCTCCACTGCCACGGTG	1240	
Db	507	TCGAGGCTGAAGAGTTCACAGACTGGTGCAGGAGAGAACTTCTCCACTGCCACGGTG	448	
Qy	1241	AGGACACTTGGTCCCTGCACGGGAGCCAGCTACTTCCATAAGACCTGGGAGCTACAG	1300	
Db	447	AGGACACTTGGTCCCTGCACGGGAGCCAGCTACTTCCATAAGACCTGGGAGCTACAG	388	
Qy	1301	ATGCTATGTCCACAGGAGGCCCCACGAAATCTCTGCCGGCTGGAGGCTGTCCAGAGGATG	1360	
Db	387	ATGCTATGTCCACAGGAGGCCCCACGAAATCTCTGCCGGCTGGAGGCTGTCCAGAGGATG	328	
Qy	1361	CTGGGGATAAGCCCTTAGGCACACAGCTTAGACACCTCCAAAGAACAGGCCCGCTGATGC	1420	
Db	327	CTGGGGATAAGCCCTTAGGCACACAGCTTAGACACCTCCAAAGAACAGGCCCGCTGATGC	268	
Qy	1421	AAGATGCCAGATCTGATACCCATTAGAGCCCCGAGAATTCCTCTTCGATCCCAAGTTTG	1480	
Db	267	AAGATGCCAGATCTGATACCCATTAGAGCCCCGAGAATTCCTCTTCGATCCCAAGTTTG	208	
Qy	1481	CAGCAAAACCCACACCCAGCTCACACAGCAAAACAAATGGACAGGCCCCAGAGGCTGAAG	1540	
Db	207	CAGCAAAACCCACACCCAGCTCACACAGCAAAACAAATGGACAGGCCCCAGAGGCTGAAG	148	
Qy	1541	CAACAGTGTCCCTCTCGCTGTGTGGAGCTCCCGAGTAACCACTATTATTTTACC	1600	
Db	147	CAACAGTGTCCCTCTCGCTGTGTGGAGCTCCCGAGTAACCACTATTATTTTACC	88	
Qy	1601	TCCTTCCCAACCTGGAGCATTTATGCTAGGCTTTGTCAAGAACTCTGTTCAGTCCCTCTC	1660	
Db	87	TCCTTCCCAACCTGGAGCATTTATGCTAGGCTTTGTCAAGAACTCTGTTCAGTCCCTCTC	28	

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

RESULT 10

AW841643

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

CONTACT

LABORATORY

LUDWIG INSTITUTE FOR CANCER RESEARCH

RUA PROF. ANTONIO PRUDENTE 109, 4 ANDAR, 01509-010, SAO PAULO-SP, BRAZIL

TEL: +55-11-2704922

FAX: +55-11-2707001

EMAIL: asimpson@ludwig.org.br

THIS SEQUENCE WAS DERIVED FROM THE FAPESP/LICR HUMAN CANCER GENOME PROJECT. THIS ENTRY CAN BE SEEN IN THE FOLLOWING URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=st2-rc1-CN0017-120

200-012-a09&t3=2000-02-12&t4=1)

Seq primer: puc 18 forward

High quality sequence start: 25

High quality sequence stop: 551.

Location/Qualifiers

1..551

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone_lib="CN0017"

/dev_stage="Adult"

/note="Organ: colon_normal; Vector: puc18; Site_1: Smal;

Site_2: SmaI; A mini-library was made by cloning products

derived from ORESTES PCR (U.S. Letters Patent application

No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the puc 18 vector. Reverse transcription of

tissue mRNA and cDNA amplification were performed under

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

low stringency conditions."

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

QY 1661 CTTCTCAATAAAAGCATCTTCAAGCTT 1687
 ||||||||||||||||||||||||||||||||
 Db 27 CTTCTCAATAAAAGCATCTTCAAGCTT 1

BASE COUNT	134 a	142 c	106 g	101 t
ORIGIN				

Query Match	27.9%;	Score 477.2;	DB 14;	Length 483;
Best Local Similarity	99.2%;	Pred. No. 5.3e-109;		
Matches 479;	Conservative	0;	Mismatches 4;	Indels 0;
Gaps 0;				

Qy 1226 TCCACTGCCACGGTGAGGACACTTGGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGA 1285

QY 1286 CCTGGGGACCTACAGATGCTATGTCCACAGGAGCCCCACGAATCCTGTCCCGGCTGGAG 1345

[illegible]

Db 121 GCTGTCAAGGATGCTGGGATGAGCCCTTAGGCACCAGCTTAGACACCTCCAGAACC 180

QY 1406 AGCCCCCGTGATGCAAGATGGAGACTGTATACCCATAGAGCCCCGAGAAATTCCTCTT 1465

QY 1466 CTGGATCCCGAGTTTGGAGCAAAACCCACACCCCAGCTCACACAGCAAAAAACAATGGACAG 1525

Db 241 CTGGATCCCAGTTTGGCAGCAAAACCCACACCCCAGCTCACACAGCAAGCAAAACAATGGACAG 300

Db 301 GCCCAGAGCTGAAGCAACAGTGTCCCTTCGGCTGTGTTGGAGCTCCCCACATAACCA 360

Qy 1586 CCTATTATTTACCTCTTTCCCAAACCTGGAGCATTTATGCCTAGGCTTGTCAGAATC 1645

22 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

Db 421 TGTTGAGTCCCTCTCCTTCTCAATAAAGCATCTTCAAGCTTGTAATAAAAAAAAAA 480

Qy	1/00	A1A	1/00
101	101	101	102

BM792701	LOCUS	BM792701	478 bp	mRNA	linear	EST 05-MAR-2002

5', mRNA sequence.
ACCESSION BM792701

KEYWORDS EST.
SOURCE human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS
Kim, N. S., Hahn, Y., Oh, J. H., Lee, J. Y., Ahn, H. Y., Chu, M. Y., Kim, M. R., Oh, K. J., Cheong, J. E., Sohn, H. Y., Kim, J. M., Park, H. S., Kim, S. and Kim, S. and

TITLE 21C Frontier Korean EST Project 2001
JOURNAL. Unpublished (2002)

CONTACT: KIM IS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology

52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4400

Email: yongsung@email.kribb.re.kr
Plate: 21 row: E column: 11

FEATURES	Location/Qualifiers
source	1. .478

```

/organism="Homo sapiens"
/db_xref="taxon:9608"
/clone="S22SNUL6n1-21-E11"
/clone_lib="S22SNUL6n1"
/sex="F"
/tissue_type="Ascites"
/cell_type="Lymphoblast-like"
/cell_line="SNU-16"
/lab_host="DH10B"
/notes="Organ: Stomach; Vector: pT73-Pac; Site_1: ECORI; Site_2: NotI; The S22SNUL6 library was contributed by the Soares laboratory and it was constructed as described by Bonaldo, M.F., Lennon, G. and Soares, M.B. (1996). Genome Research 6(9): 791-806. RNA was prepared from harvested cells of SNU-16 culture. SNU-16 cell was obtained from Korean Cell Line Bank (KCLB). SNU-16 was established from ascitic fluids of Korean patients by Park J.G. et al. (1990). Cancer Res 50: 2773-2780."
123 a 143 c 109 g 103 t

```

Query Match	27.7%;	Score 473.2;	DB 14;	Length 478;
Best Local Similarity	99.4%.			
Prod NO	5	30-108.		

[illegible]

Db 1 CAGAGAACTTCTCCACTGCCACGGTGAGGACACTTGGGTCCCTGATGGGGAGCCAGCTAC 60

QY 12/5 TTTCGCATAAAGACCTGGGGAGCTACAGATGCTATGTTCCACAGGAGGGCCCCACCGAATCTCTGT 1334
=====

QY 1335 CCCGGCTGGAGGTGTCAGAAGGATGCTGGGGATAAGCCCTTAGGCACCAGCTTAGACAC 1394

Db 121 CCCGGTGGAGGCTGTCAGAAGGATGCTGGGGATAAGCCCTTAGGCACCAGCTAGACAC 180

Db 181 CTCCAGAACCAGGCCCTGATGCAAGATGTCAGATCTGATACCATTAGAGCCCCGA 240

Qy 1455 GAATTCCTCTTCTGGATCCAGTTTTCGACGCAAAACCCACACCCAGTCTACACAGCAAAA 1514

1515 ACAATGGACAGGCCCAGAGGCTGAAGCAACAGTGTCCCTTCTGGCTGTGTTGGAGGCCNC 1574

Db 301 ACAATGGACAGGCCCCAGAGGGTGAAGCAACAGTGTCCCTTCTGGCTGTGTTGGAGCCTC 360

QY	CCCCAGTATCCACGCTATTATTTATTCCTC	TCCCCTATCGCTGGGTGCAATTAATGCCATACCGC	1054
Dh	CCCCAGTATCCACGCTATTATTTATTCCTC	TCCCCTATCGCTGGGTGCAATTAATGCCATACCGC	420

Qy 1635 TGTCAAGAAATCTGTTTCAGTCCCTCTCCTTCTCAATAAAAGCATCTTCAAGCTTGTTAA 1692

Db 421 TGTCAGAAATCTGTTTCAGTCCCTCTCCTTCTCAATAAAGCAATCTTCAAGCTTGAAAA 4/8

RESULT 13
AW937714/C

DEFINITION QV3-DT0045-221299-046-a01 DT0045 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW937714

KEYWORDS EST, SOURCE human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Naraj M A, da Silva W Jr, Zago M A, Bordin S, Costa F F

601010000, G.N.: Caratteristiche, V.F.T.;
MacSweeney, W.A.: Data, G.C.;
Bosch, J.P.: B.I.

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=&t2=QV3-DT0045-221
299-046-a01&t3=1999-12-22&t4=1)
299-046-a01&t3=1999-12-22&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 16
High quality sequence stop: 365.

FEATURES

source

1. .534
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="DT0045"
/dev_stage="Adult"

/note="Organ: denis_drash; Vector: puc18; Site1: SmaI;
Site2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 96 a 143 c 162 g 133 t

ORIGIN

Query Match 27.6%; Score 472; DB 10; Length 534;
Best Local Similarity 97.2%; Pred. No. 1.1e-107;
Matches 513; Conservative 0; Mismatches 10; Indels 5; Gaps 3;
QY 249 CAAGTGCTTTACTGCTTCTCCAGGCGGTCCTTCCCGGAGGAGGAGGAGG 308
DB 534 CAAGTGCTTACTGCTTCTCCAGGCGGTCCTTCCCGGAGGAGGAGGAGG 475
QY 309 ACAGGAGTGTCTCAACCATGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 368
DB 474 ACAGGAGTGTCTCAACCATGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 415
QY 369 AGGCCAGGCAAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 425
DB 414 AGGCCAGGCAAGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 355
QY 426 GAGGTCTCACCAGGACAGATACA-TTGACTGCTTCCAGAGATCAAGTACAGCTTCAAC 484
DB 354 GAGGTCTCACCAGGACAGATACA-TTGACTGCTTCCAGAGATCAAGTACAGCTTCAAC 295
QY 485 CTCTGGGAAGGCTGGCCACCTGGCTCAAGGAGACAGTCCCTGAGCTCGTACACATC 544
DB 294 CTCTGGGAAGGCTGGCCACCTGGCTCAAGGAGACAGTCCCTGAGCTCGTACACATC 235
QY 545 CTCTTCAAGTCCCTGAACTTCATCTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 604
DB 234 CTCTTCAAGTCCCTGAACTTCATCTCCGCGGAGGAGGAGGAGGAGGAGGAGGAGG 175
QY 605 GTGATCTCACCCTCTCTCACCCTTAAAGCTATCAAGCTGTACAGTCTCTGTCTAAGCCCA 664
DB 174 GTGATCTCACCCTCTCTCACCCTTAAAGCTATCAAGCTGTACAGTCTCTGTCTAAGCTCA 115
QY 665 CTTGAGAGTAACTTTGGATGGGGTGGGGCCAGGCTGGGAGGAGGAGGAGGAGGAGGAGG 724
DB 114 CTTGAGAGTAACTTTGGATGGGGTGGGGCCAGGCTGGGAGGAGGAGGAGGAGGAGGAGG 55

QY 725 ACAGGGCATGAGCCCTGCGCTTACCAACCCACATTTCTCAGATGACTGG 772
DB 54 ACAGGGCATGAG-CCCTGCGCTTACCAACCCACATTTCTCAGACCGCAGGG 8
RESULT 14
AI339104/c
LOCUS
DEFINITION
AI339104
VERSION
AI339104.1 GI:4076031
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 456)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D.
CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Insert Length: 606 Std Error: 0.00
Seq primer: -40UP from Glbco
High quality sequence stop: 447.
Location/Qualifiers
1. .456
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:1946780"
/clone_lib="NCI-CGAP_GC4"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
polylinker: pT73D-Pac (Pharmacia) with a modified
germ cell tumors, and was then primed with a Not I -
oligo(drf) primer. Double-stranded cDNA was ligated to Eco
RI adaptors (Pharmacia), digested with Not I and cloned
into the Not I and Eco RI sites of the modified pT73
vector. Library is normalized. Library was constructed by
Bento Soares and M. Fatima Bonaldo."
BASE COUNT 98 a 105 c 137 g 116 t
ORIGIN
Query Match 26.6%; Score 454.4; DB 9; Length 456;
Best Local Similarity 99.8%; Pred. No. 2.8e-103;
Matches 455; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1236 CGGTGAGGACATTTGGTCCCTGACGGGAGGAGGAGGAGGAGGAGGAGGAGG 1395
DB 456 CGGTGAGGACATTTGGTCCCTGACGGGAGGAGGAGGAGGAGGAGGAGGAGG 397
QY 1296 TACAGATGCTATCTCCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1355
DB 396 TACAGATGCTATCTCCACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 337
QY 1356 GGATGCTGGGGATAAGCCCTTAGGCACACAGCTTTAGACACCTCCCAAGAACAGCCCGCT 1415
DB 336 GGATGCTGGGGATAAGCCCTTAGGCACACAGCTTTAGACACCTCCCAAGAACAGCCCGCT 277
QY 1416 GATTCAGATGGGAGATCTGATACCATTTAGACCCCGGAGAGATTCCTTCTTGATCCCCA 1475

Db 276 GATGAAGATGGCAGATCTGTATACCCATTAGAGCCCCGAGAAATTCCTTCTGTGATCCCA 217

QY 1476 GTTTCAGCAAAACCCACACCCAGCTCACACAGCAAAACAATGAGAGGCCAGAGGC 1535

Db 216 GTTTCAGCAAAACCCACACCCAGCTCACACAGCAAAACAATGAGAGGCCAGAGGC 157

QY 1536 TGAAGCAAAACAGTGCCTTCTGGCTGTGTGGAGCCCTCCCGAGTAACCACTATTATT 1595

Db 156 TGAAGCAAAACAGTGCCTTCTGGCTGTGTGGAGCCCTCCCGAGTAACCACTATTATT 97

QY 1596 TTACCTCTTCCCAAACTGGAGCAATTTATGCTAGGCTTGTCAAGATCTGTTCACTCC 1655

Db 96 TTACCTCTTCCCAAACTGGAGCAATTTATGCTAGGCTTGTCAAGATCTGTTCACTCC 37

QY 1656 CTCTCTCTCTCAATAAAAGCATCTTCAAGCTTGTA 1691

Db 36 CTCTCTCTCTCAATAAAAGCATCTTCAAGCTTGAAA 1

RESULT 15

BO951422

LOCUS

DEFINITION AGENCOURT_8836792 NCI_CGAP_Co24 Mus musculus cDNA clone

IMAGE:6397884 5', mRNA sequence.

BO951422

BO951422.1 GI:22365900

EST.

SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.
Email: cdapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Plate: LLAM13896 row: h column: 13
High quality sequence stop: 651.

FEATURES

source

Location/Qualifiers

1..901

/organism="Mus musculus"

/strain="FVB/N"

/db_xref="taxon:10090"

/clone="IMAGE:6397884"

/clone_lib="NCI_CGAP_Co24"

/lab_host="DH10B (T1 phage-resistant)"

/note="Organ: colon; Vector: pCMV-SPORT6; Site:1; Noti; Site:2; Sali; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.6 kb. Constructed by Life Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT 247 a 246 c 225 g 180 t 3 others

ORIGIN

Query Match 25.9%; Score 443.6; DB 14; Length 901;

Best Local Similarity 71.4%; Pred. No. 1.7e-100;

Matches 640; Conservative 0; Mismatches 243; Indels 13; Gaps 4;

QY 816 ACCCTGTTTCCCTCGCGGGGAAGTCATAGTTAGGAGCACTTCCACTTTCCTCAGG 875

Db 15 ACCAGGATTCGATTTCCTCCGAGGTTCTAGGATGAGGAGCGCTACATTTCCCTCGG 74

QY 876 AGAAGACACACACCACTGACCTAGCCTGGGGCCCCCAACTCCAGGCCCTCCAGCCCA 935

Db 75 ATGAGCCATACAAACCAATACCCCTGAATACGAGGACTCAAAACCTCCACTCTCCAGCCCA 134

QY 936 AACCTGCCAGCCAGCCCTGAAATGCAAGTCTTGTACAGTTTGAAGCTAGGAACCCAC 995

Db 135 GCCCTGGCAGAGCAGCCCTGAAATGCAAGTCTTGTACAGTTTGAAGCAAGATGCC 194

QY 996 GGGAACTGACTGTGTGTCAGGGAGAGAGCTGGAGTTCTGGACCAACACAGCGGTGT 1055

Db 195 AGGAACTGACTGTGCACAGGGGGAGATTCTGGAGTTCTGGATCAGACAGCGATGT 254

QY 1056 GGCTGTTGAAGAAATCAGCGGGGAGGCTACATTCCTCAAGCAACATCTCTGGAGCC 1115

Db 235 GGCTGTTGAAGAAATCAGCGGGGAGGCTACATTCCTCAAGCAACATCTCTGGAGCC 314

QY 1116 TACAGCCGGGGACCCCTGGGACCCAGGGCCAGTCAACCTCTCTGGGTTCCAAATGCTTCGAC 1175

Db 315 TGCCAGCTGGAGCCCGGAGAGGCCACCGACAACCATCTTTAGGGCCCCCAATGCTTCGCC 374

QY 1176 TTAGCTCGAGGCTGAAGAGGTCACAGAGTGGCTGCAGGCAGAGAACTTCTCCACTGCCA 1235

Db 375 TGAGCTCAAGCTTGAGGAGGTCACAGCTGGCTTCAAGCAGAGAACTTCTCCACTGTCA 434

QY 1236 CGGTGAGGACACTTGGTCCCTGACGGGAGGAGCTACTTTCGATAAGACCTTGGGGAGC 1295

Db 435 CTGTAGGAGCCCTCGGTCCTTAATGGGAGCCAGCTACTTCAATGAGACCTGGGGAGC 494

QY 1296 TACAGATGCTATGTCCACAGAGGCCCCACGAAATCTGTCCCGGCTGGAGGCTGTGAGAA 1355

Db 495 TGCAGATGCTGTGTCCACAGAGGAGCTCCAGGGTCCAGGCCCGCTGGAGCTGTGAGAA 554

QY 1356 GGATGCTGGGATAGCCCTTTAGGCACAGCTTTAGACACCTTCAAGAACCCAGG-CCCCGC 1414

Db 555 GGATGCTGGGATGACTCATTTAGAGACCCAGCTCGAATACCTCTAAGACCAAGGACCTCTC 614

QY 1415 TGATCAAGATGGCAGATCTGTATACCCA-TTAGAGCCCCGAGAAATTCCTTCTTGGATCC 1473

Db 615 ATGTCAAGATGGAGTATTTGGTACCCAGTTAGAGCCCTTGGAAGTCTCTT-----T 667

QY 1474 CAGTTTGCAGCAAAACCCACACCCAGCTCACACAGCAAAACAATGGAGGCCAGAG 1533

Db 668 CAGTTAGTGGCAAGCCACATATCTTGGATCAACAAGCANAGAAGATGAACAGTCCAGAC 727

QY 1534 GCTGAAGCAAAACAGTGTCCCTTCTGTGCTGTGTGGAGCTCCCGAGTAACCACTATTATA 1593

Db 728 ACTGAGGCAAAATGGTACCTCATTTGGG---GTTGGAGACCTCACCGGGGTACTATTATA 783

QY 1594 TTTTACCTCTTTCCCAAACTCGAGCATTTATGCCCTTAGCTTGTCAAGAAATCTGTTCACT 1653

Db 784 TTTACCCCAACAACCTTAAGTGGCATATTATTTCCCGAGGCTACTCCGGATCTCTTTAGT 843

QY 1654 CCCTCTCTCTTCAATAAAAGCATCTTCAAGCTTGTAAATAAAAAAATAANGATAA 1709

Db 844 CCCAATCTGGCAATAAAAAACATCGGTTTAACCTGGTCTTAACAAAAAATAAAAAA 899

Search completed: February 25, 2003, 16:09:44
Job time : 1702 secs

THIS PAGE BLANK (USPTO)


```
Db 219 ValProGlyThrValThrGlnValAspValArgSerArgValAlaAlaTrpSerAla 238
QY 242 CGAGAACCAAGTGCCTTACTCTGCTCCCAAGCGGTCTCTTCCCGGAGGAC--- 298
Db 239 TrpAlaAlaAspGlnGlyAspPheGluLysProArgGlnTyrHisGluGlnGluThr 258
QY 299 CCAGAG-----AGGACGAGGAAGTCTGCAACATGCTCTCAAGG 337
Db 259 ProGluMetMetAlaAlaArgIleAspArgAspValGlnIleLeuAsnHisIleLeuAsp 278
QY 338 GACATTGAGCTTTCATGGGAAGCTGGAGAGGCCAG-----CCA 379
Db 279 AspIleGluPhePheIleThrLysLeuGlnLysAlaAlaGluAlaPheSerGluLeuSer 298
QY 380 AGACACGACGAGCAAGAAATTTGGGAAAAAACAAGGACCGAGGAGTCTCACC--- 436
Db 299 LysArgLysLysSerLysLysSerLysArgLysGlyProGlyGluGlyValLeuThrLeu 318
QY 437 -----CAGGCACAGTACATTGCTTCCAGAAAGTCAAGTACAGC 478
Db 319 ArgAlaLysProProProProAspGluPheValAspCysPheGlnLysPheLysHisGly 338
QY 479 TTCACCTCTGGGAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCTCAGCTGTA 538
Db 339 PheAsnLeuLeuAlaLysLysSerHisIleGlnAsnProSerAlaSerAspLeuVal 358
QY 539 CACATCCTCTCAAGTCCCTGAACCTTCATCCTGGCCAGGTGCCCTCAGGCTGGCCTAGCA 598
Db 359 HisPheLeuPheThrProLeuAsnMetValValGlnAlaThrGlyGlyProGluLeuAla 378
QY 599 GCCCAAGTGTATCATCCCTCTCACCCCTAAAGCTATCAACCTGCTCAGTCTGTCTA 658
Db 379 SerSerValLeuSerProLeuLeuThrLysAspThrValAspPheLeuAsnTyrThrAla 398
QY 659 AGCCCACTGAGAGTAACTTTGGATGGGTGGGCCCGCCAGCTGACCACTAGCGGGCC 718
Db 399 ThrAlaGluGluArgLysLeuTrpMetSerLeuLeuLysAspSerTrpValLysValArgAla 418
QY 719 GACTGGACAGGGGATGAGCCCTG---CCCTACCAACCCACATCTTCAGATCAGTGGCAA 775
Db 419 GluTrpProLysGluGlnPheIleProTyrValProArgPheArgAsnGlyTrpGlu 438
QY 776 CTTCCA-----GAGCCCTCCAGCCCAAGCACCCCTTAGGATACCGAGAC 817
Db 439 ProProMetLeuAsnPheMetGlyAlaProThrGluGlnAspMet-----TyrGlnLeu 456
QY 818 CCTGTTTCCCTTCGGCGGGAGTCAAT---AGTTAGGAGGACACCTCAGCTTCTCTCAG 874
Db 457 AlaGluSerValAlaAsnAlaGluHisGlnArgLysGlnAspSerLysArgLeuSerThr 476
QY 875 GAGAAGACACACAAACCATGACCTCAGCCTGCGGAG----- 910
Db 477 GluHisSerAsnValSerAspTyrProProAlaAspGlyTyrAlaTyrSerSerMet 496
QY 911 -----CCC 913
Db 497 TyrHisArgGlyProHisAlaAspHisGlyGluAlaAlaMetProPheLysSerThrPro 516
QY 914 AAC-----TCCAGGCGCTCCAGCCCAACCTGCGCCAGCCCTGAAATGCAAA 964
Db 517 AsnHisGlnValAspArgAsnTyrAspAlaValLysThrGlnProLysLysTyrAlaLys 536
QY 965 GTCTTGTACGAGTTTGAAGTACGAGCCAGCGGAACCTGCTGTGTCCAGGAGAGAG 1024
Db 537 SerLysTyrAspPheValAlaArgAsnSerSerGluLeuSerValMetLysAspAspVal 556
QY 1025 CTGAGGTTCTGACCAACGAGCGGTGGTGGTGGTGAAGATGAGCGGAGCGGAC 1084
Db 557 LeuGluIleLeuAspAspArgArgGlnTrpTrpLysValArgAsnAlaSerGlyAspSer 576
QY 1085 GGCTACATTCACAGCAACATCCTGGAG----- 1111
Db 577 GlyPheValProAsnAsnIleLeuAspIleMetArgThrProGluSerGlyValGlyArg 596
```

```
QY 1111 ----- 1111
Db 597 AlaAspProTyrThrHisThrIleGlnLysGlnArgThrGluTyrGlyLeuArgSer 616
QY 1112 -----CCCCTA 1117
Db 617 AlaAspThrProSerAlaProSerProProProThrProAlaProValProValProLeu 636
QY 1118 CAGCGGGGACCCCTGGG----- 1135
Db 637 ProProSerValProAlaProValSerValProLysValProAlaAspValThrArgGln 656
QY 1135 ----- 1135
Db 657 AsnSerSerSerAspSerGlyGlySerIleValArgAspSerGlnArgTyrLysGln 676
QY 1135 ----- 1135
Db 677 LeuProValAspArgArgLysSerGlnMetGluGluValGlnAspGluLeuPheGlnArg 696
QY 1136 ---ACCCAGGCGCAGTCAACCTCTCGG-----GTTCCA 1165
Db 697 LeuThrIleGlyArgSerAlaAlaGlnArgLysPheHisValProArgGlnAsnValPro 716
QY 1166 ATGCTTTCGACTTAGC-----TCGAGGCGCTGAAGAGTGCACAGACTGGCTGCAGCAGAG 1219
Db 717 ValIleAsnIleThrTyrAspSerSerProGluGluValLysThrTrpLeuGlnSerLys 736
QY 1220 AACTTCTCCACTGCCCGTGCAGGACACTGGTCCCTGCAGCGGGAGCCAGCTACTTCCG 1279
Db 737 GlyPheAsnProValThrValAsnSerLeuGlyValLeuAsnGlyAlaGlnLeuPheSer 756
QY 1280 ATAACACCTGGGAGCTACAGATGCTATGTCCACAGGAGGCCACCAATCCTGTCGCGG 1339
Db 757 LeuAsnLysAspGluLeuArgSerValCysPro---GluGlyAlaArgValPheAsnGln 775
QY 1340 CTGGAGGCTGTCTCAGAGGATGCTG 1363
Db 776 IleThrValGlnLysAlaAlaLeu 783
```

RESULT 2
138728
epidermal growth factor receptor kinase substrate - human
C:Species: Homo sapiens (man)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 09-Jun-2000
C:Accession: I38728
R:Wong, W.T.; Carlomagno, F.; Druck, T.; Barletta, C.; Croce, C.M.; Huebner, K.; Krau
Oncogene 9, 3057-3061, 1994.
A:Title: Evolutionary conservation of the EP88 gene and its mapping to human chromoso
A:Reference number: I38728; MUID:94366758; PMID:8084614
A:Accession: I38728
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-822 <RES>
A:Cross-references: EMBL:U12535; NID:G530822; PID:G530823
C:Genetics:
A:Gene: Eps8
C:Superfamily: SH3 homology

Alignment Scores:
Pred. No.: 7,33e-31 Length: 822
Score: 550.00 Matches: 158
Percent Similarity: 38.71% Conservative: 77
Best Local Similarity: 26.03% Mismatches: 174
Query Match: 17.59% Indels: 198
DB: 2 Gaps: 15

US-09-762-021A-1 (1-1710) x I38728 (1-822)

QY 101 GGCCTGTCTATGAAAGCGCGCTCCCTATGTGAGCAGCAGCTATCTGTGAGCGCGGATC 160
Db 190 GlyLysGlnLysArgArgProAspAlaLeuArgMetIleSerAsnAlaAspProSerIle 209

Qy	1028	GAGTTCCTGGACACACACGCGTGGTGGCTGCTCAACAANTAGCGGGACGGACGGC	108
Db	559	GlutLeuAspArgLysGlnTrpTrpLysValArgAsnAlaSerGlyAspSerGly	578
Qy	1088	TACATTCCAAAGCAACATCCCTGGAG-----	1111
Db	579	PheValProAsnAsnLleLeuAspLleValArgProProGluSerGlyLeuGlyArgAla	598
Qy	1111	-----	1111
Db	599	AspProProTyrThrHisThrIleGlnLysGlnArgMetGluTyrGlyProArgProAla	618
Qy	1112	-----CCCTACAG	1122
Db	619	AspThrProProAlaProSerProProThrProAlaProValProValProLeuPro	638
Qy	1121	CGGGGACCCCTGGG-----	1133
Db	639	ProSerThrProAlaProValProValSerLysValProAlaAsnIleThrArgGlnAsn	658
Qy	1135	-----	1135
Db	659	SerSerSerSerAspSerGlyGlySerIleValArgAspSerGlnArgHisLysGlnLeu	678
Qy	1135	-----	1135
Db	679	ProValAspArgArgLysSerGlnMetGluValGlnAspGluLeuIleHisArgLeu	698
Qy	1136	ACCCAGGCGCAGTACACCTCTCGG-----	1166
Db	699	ThrIleGlyArgSerAlaAlaGlnLysLysPheHisValProArgGlnAsnValProVal	718
Qy	1169	CTTCGACTTACG-----TCGAGGCGCTGAAGAGGTCCACAGCTGGCTGCAGCAGAGAAC	1222
Db	719	IleAsnIleThrTyrAspSerThrProGluAspValLysThrTrpLeuGlnSerLysGly	738
Qy	1223	TTCTCCACTCCGCGGTGAGGACACATTGGTCCCTGACGGGAGCCAGCTACTTCGCAT	1288
Db	739	PheAsnProValThrValAsnSerLeuGlyValLeuAsnGlyAlaGlnLeuPheSerLeu	758
Qy	1283	AGACCTGGGGAGCTACAGATGCTATGCCAGAGAGCCCGCCAGCAATCTGCCCGGCTG	1344
Db	759	AsnLysAspGluLeuArgThrValCysPro--GluGlyAlaArgValTyrSerGlnIle	777
Qy	1343	GAGCGCTGCAGAAGGATGCTG	1363
Db	778	ThrValGlnLysAlaAlaLeu	784
RESULT 3			
T27237			
hypothetical protein Y57G11C.24c - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999			
C:Accession: T27237			
R:McMurray, A.			
submitted to the EMBL Data Library, September 1997			
A:Reference number: Z20330			
A:Accession: T27237			
A:Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-929 <MIL>			
A:Cross-references: EMBL:Z99281; PIDN:CAB16526.1; GSPDB:GN00022; CESP:Y57G11C			
A:Experimental source: clone Y57G11C			
C:Genetics:			
A:Gene: CESP:Y57G11C.24c			
A:Map position: 4			
A:Introns: 17/2; 61/1; 134/3; 210/3; 252/1; 355/3; 433/2; 538/3; 575/3; 686/3			
Alignment Scores:			
Pred. No.:	3,31e-17	Length:	929
Score:	356.00	Matches:	145
Percent Similarity:	39.21%	Conservative:	84
Best Local Similarity:	24.83%	Mismatches:	209

Alignment scores:	
Pred. No.:	3,31e-17
Score:	356.00
Percent Similarity:	39.21%
Best Local Similarity:	24.83%
Length:	929
Matches:	145
Conservative:	84
Mismatches:	209

Query Match: 11.38% Indels: 146
DB: 2 Gaps: 23
US-09-762-021a-1 (1-1710) x T27237 (1-929)
QY 224 CCAGCGCACACGAGTGGCGGAGAACCAAGTGCCTTACTCTGCTCCCTCCCAAGGGGGTCC 283
DB 243 ProThrHisLeuGlnAlaGlnGlnMetProPheThrProProAspAlaSer 262
QY 284 TCTTCCCCGAGGACCCA-----GAGAGGAGCAGGAGTGTGAACCATGTCCTA 334
DB 263 IleSerSerGluThrSerGluMetPheGluArgAspValAsnThrLeuAsnArgCysPhe 282
QY 335 AGGACATTGAGCTGTTCATGGGAAGCTGGGAAGGCC----- 373
DB 283 AspAspIleGluArgPheValAlaArgIleGlnSerAlaAlaLeuAlaGlnArgGluIle 302
QY 374 -----CAGCAAGACCCAGCAGAGAGAGAAATTTGGGAAAAAAC 415
DB 303 GluGlnGlnAsnHisArgTyrArgThrAlaAsnArgArgAspLysLysAsnGlnGlnPro 322
QY 416 AGGACACGAGGAGCTC-----ACCCAGGCACAGTACATT 451
DB 323 ProAspProAsnGlyIleLeuPheMetArgAlaGlnLeuProLeuGluSerGluPheVal 342
QY 452 GACTGCTTCCAGAAGATCAAGTACAGCTTCAACCTCTCCGGAAGGCTGCCACCTGGGCTG 511
DB 343 AspIleLeuLysLysPheLysLeuSerPheAsnLeuLeuAlaLysLeuLysAsnHisIle 362
QY 512 AGGAGACAAAGTGGCCCTGAGCTGTACATCTCTTCAAGTCCCTGAACCTCATCTG 571
DB 363 HisGluProAsnAlaProGluLeuLeuHisPheLeuPheThrProLeuAlaValIleLeu 382
QY 572 GCCAGTGCCT-----GAGCTGGCTAGCAGCCCAAGTATCTACCCCTCCTC 622
DB 383 GluAlaCysHisTrpGlyLeuGlyArgAsnValAlaProThrValAlaSerProLeuLeu 402
QY 623 ACCCTTAAGCTATCAACCTGTACAGTCTCTTAAAGCCCACTGAGAGTAACTTTGG 682
DB 403 SerLeuGluAlaArgGluLeuMetGlnAsnCysLeuThrSerHisGluSerAspIleTrp 422
QY 683 ATGGGTGGGCGCCAGCTGGACCACT-----AGCGGGCGGCTGGACAGCGCAT 733
DB 423 MetSerLeuGlyGluAlaTrpArgThrProProAspGluLysArgArgMethHisAlaGlu 442
QY 734 GAGCCCTCCCTACCAACCCCACTTCTCAGATGACTGGCAACTCCAGAGCCCTCC--- 790
DB 443 LysAspLeuIleThrLysGluThrThrGln-----ProValProProAlaAla 459
QY 791 -----AGCCAAGCACCCCTTA-----GGATACAGGACCCCTGTTTCCCTCGCGGGGA 838
DB 460 ValValThrHisGlnProIleThrLysArgTyrAspProProIleSerIle----- 476
QY 839 AGTCATAGTTAGGAGCACCTCACACTTCTCAGGAGAGACACACCAACCATGACCT 898
DB 477 -----SerProProGlnArgAsnAsnTyrSerHisValLys 489
QY 899 CAGCCTGGGACCCCACTCCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTG--- 955
DB 490 ValThrValAsp-----SerAspThrSerProArgGlnGlnAlaPheIle 504
QY 956 -----AAATGCAAGTCTTG---TAGCAGTTTGAAGCTAGG 988
DB 505 AspAspIleValAlaLysGlyGlyLysLeuAlaValValThrTyrAspArgGlyGlyGln 524
QY 989 ACCCCAGGGAAGTACTGTGTCTCAGGAGAGAGAGAGTGTCTGGACACACAGCAG 1048
DB 525 AsnThrLysGluLeuThrValHisLysGlyGlyGlyGlyGlyGlyGlyGlyGlyGly 544
QY 1049 CGTGTGTGCTGGTGAAGATGAGCGGGGAGCGGCGGTACATTCACCAAGCAACATCTG 1108
DB 545 AsnTrpTrpGluCysLysAsnMethHisGlnArgValGlyTyrValProHisThrIleLeu 564

QY 1109 GAGCCCCCTACAGCCGGGACCCCTGGGACCCAGGCGCCCTACCTCTCGGGTCCAATG 1168
DB 565 SerMetVal-----ProPheGluGlnGlnTyrAlaGlnGlnTyrAsnVal 580
QY 1169 CTTCCAGCTAGCTCAGAGCCTGAGAGAGTACAGACTGGCTGGTCCAGGCAGAGA---GAACTT- 1224
DB 581 MetSerSerGlnHisAsnValIleGlnIleAsnAsp--GlyValGlyLysMetGluValP 600
QY 1225 -----CTCCACTGCCACGGTG 1240
DB 600 heIleSerLysThrCysAsnPheIleLeuIleThrThrCysAsnLeuPheCysSerThrT 620
QY 1241 AGGACACTTGGTCCCTGCAC----- 1260
DB 620 hrLysThrAlaValProAspIleTyrArgAlaPheValMetAsnAspGluThrArgGlnM 640
QY 1261 -----GGGAGCCACTACTTCGCATACACCTGGGGAGCTACAGATGCTATGTCACAG 1315
DB 640 etSerGlyThrProCysGluAsnSerArgSerPheLeuSerSerAsnThrLeuSerSer- 659
QY 1316 GAGGCCCCCAAGTCTCTCCCGCTGGAGGCTGTCAAGAGGATCTGGGGATAAGCCCT 1375
DB 660 ---SerProGluAsn-----AlaProGluIleProL 669
QY 1376 TAGGCACCACTTAGACACCTCCAAAGAACCCAGCCCGCTGATGCAAGATGCGAGATCTG 1435
DB 669 euLeuGluAsnLeuLysPheAsnHisAsnLysLeuTyrLeuPheHisIleLeuSerProA 689
QY 1436 ATACCATTAGAGCCCGAGAAATTCCTCTCT- 1467
DB 689 snHisHis-----AsnAsnSerSerSerThrGlyGlyTyrAsnAsnGlyHisHisG 706
QY 1468 --GGATCCCACTTTCAGCAAAACCCACACCCAGCTCACACAGCAAAACAAATGGACAG 1525
DB 706 InGlyProGlyMetLeuProGluAspAlaProSerTyrValLysGluArgGln----- 723
QY 1526 GCCCAGAGGCTGAAGCAACACAGTGC-----CCTTCTGGCTGTGTGGAGCCT 1573
DB 724 -----AlaGlnAsnValTyrArgProProProProLeuValSerAspA 739
QY 1574 CCCCACTAACCACTATTTA-----TTTTACC 1600
DB 739 IaGlyValGlnValGluIleArgArgGluHisValAlaProProProProValValI 759
QY 1601 TCTTTCCTCCAACTGGAGCATTTATGCTAGGCTTGTCAAGAAATCTGTTCAGTCCCTCTC 1660
DB 759 leProProProProProProArgLysProProValGluProValTyrGlnProGlnP 779
QY 1661 CTTCTCAA 1668
DB 779 toAlaGln 781
RESULT 4
T27238
hypothetical protein Y57G11C.24a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T27238
R:McMurray, A.
submitted to the EMBL Data Library, September 1997
A:Reference number: Z20330
A:Accession: T27238
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-732 <=IIL>
A:Cross-references: EMBL:Z99281; PIDN:CAB16527.1; GSPDB:GN00022; CESP:Y57G11C.24a
A:Experimental source: clone Y57G11C
C:Genetics:
A:Gene: CESP:Y57G11C.24a
A:Map position: 4
A:Introns: 17/2; 61/1; 134/3; 210/3; 252/1; 355/3; 433/2; 538/3; 575/3; 686/3
Alignment Scores:


```
Pred. No.: 6.21e-17 Length: 732
Score: 352.00 Matches: 134
Percent Similarity: 40.27% Conservative: 75
Best Local Similarity: 25.82% Mismatches: 186
Query Match: 11.26% Indels: 124
DB: 2 Gaps: 20

US-09-762-021A-1 (1-1710) x T27238 (1-732)

Qy 224 CCAGCCACACAGTGCAGGACCAAGTCCCTTACTCTGCCTCTCCCTCAAGGGGTCC 283
Db 224 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 243 ProThrHisLeuGlnAlaGlnGlnMetProPheTyrProPheAspAlaSer 262
Qy 284 TCTTCCCCCGAGGCCA-----GAGAGGACGAGGAGTCTGAACCATGTCTCTA 334
Db 284 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 263 IleSerSerGluThrSerGluMetPheGluArgAspValAsnThrLeuAsnArgCysPhe 282
Qy 335 AGGACATTTAGCTTTCATCGGGAAGCTGGAGAGGCC----- 373
Db 335 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 283 AspAspIleGluArgPheValAlaArgIleGlnSerAlaLeuAlaGlnArgGluIle 302
Qy 374 -----CAGGCAAGACACGACGAGGAAGAATAATTGGGAAAAAACAAC 415
Db 374 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 303 GluGlnGlnAsnHisArgTyrArgThrAlaAsnArgArgAspLysLysAsnGlnPro 322
Qy 416 AAGGACACGAGGAGTCTC-----ACCGAGGCACAGTACATT 451
Db 416 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 323 ProAspProAsnGlyIleLeuPheMetArgAlaGlnLeuProLeuGluSerGluPheVal 342
Qy 452 GACTGCTTCCAGAGATCAAGTACAGTTCACCTCCTCGGAAGCTGGCCACCTGGCTG 511
Db 452 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 343 AspIleLeuLysPheLysLeuSerPheAsnLeuAlaLysLeuLysAsnHisIle 362
Qy 512 AAGGACACAGTGGCCCTGAGCTCGTACACATCTCTTCAAGTCCCTCAATCTCATCTG 571
Db 512 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 363 HisGluProAsnAlaProGluLeuLeuHisPheLeuThrProLeuAlaValIleLeu 382
Qy 572 GCCAGGTGCCCT-----GAGGCTGGCCTAGCAGCCCAAGTATCTACCCCTCTCT 622
Db 572 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 383 GluAlaCysHisTrpGlyLeuGlyArgAsnValAlaProThrValAlaSerProLeuLeu 402
Qy 623 ACCCTTAAGCTATCAACCTCTACAGTCTGTCTAAGCCACCTGAGAGTAACCTTTGG 682
Db 623 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 403 SerLeuGluAlaArgGluLeuMetGlnAsnCysLeuThrSerHisGluSerAspIleTrp 422
Qy 683 ATGGGTTGGCCCGCTGACCACT-----ACCGCGCGCAGCTGGACAGGCAT 733
Db 683 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 423 MetSerLeuGlyGluAlaTrpArgThrProProAspGluLysArgArgMetHisAlaGlu 442
Qy 734 GAGCCCTTGCCTACCAACCCACATCTCTCAGATGACTGGCAACTTCCAGAGCCCTCC 790
Db 734 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 443 LysAspLeuIleThrLysGluThrGln-----ProValProProAlaAla 459
Qy 791 -----ACCGAAGCACCTTA-----GGATACAGGACCTGTTCCTTCCCTTCGGCGGGA 838
Db 791 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 460 ValValThrHisGlnProIleThrLysArgTyrAspProIleSerIle----- 476
Qy 839 AGTCATAGTTAGGAGACACCTCACATTTCTCAGGAGAGACACACAACCATGACCT 898
Db 839 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 477 -----SerProProGlnArgAsnAntySerHisValLys 489
Qy 899 CAGCCTGGGGACCCCACTCCAGCCCTCCAGCCCAACCTGCCACCGACCGCTG--- 955
Db 899 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 490 ValThrValAsp-----SerAspThrSerProArgGlnGlnAlaPheIle 504
Qy 956 -----AAATGCAAGTCTTG-----TACGAGTTTGAAGCTAGG 988
Db 956 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 505 AspAspIleValAlaLysGlyGlyLysLeuAlaValThrTyrAspArgGlyGlyGln 524
Qy 989 AACCCCGGGAACCTGCTGGTCCAGGAGAGAGCTGGAGGTCTCGGACCAACGACG 1048
Db 989 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 525 AsnThrLysGluLeuThrValHisLysGlyGlyGlyLeuGluValIlePheAspGluArg 544
```

```
Qy 1049 CGGTGGTGGCTGGTGAAGAATGAGCGGGACGAGCGGCTACATTTCCAAAGCAATCCTG 1108
Db 545 AsnTrpTrpGluCysLysAsnMetHisGlnArgValGlyTyrValProHisThrIleLeu 564
Qy 1109 GAGCCCTACAGCCGGGACCCCTGGGACCCAGGCCAGTCCACCTCTCGGGTTCCAATG 1168
Db 565 SerMetVal-----ProPheGluGlnGlnGlnTyrAlaGlnGlnTyrAsnVal 580
Qy 1169 CTTGACTTAGTCTGAGGCCTGAAGAGCTCACAGCTGGCTGCAGGAGAGA---GAACTT- 1224
Db 581 MetSerSerGlnHisAsnValIleGlnIleAsnAsp---GlyValGlyLysMetGluValP 600
Qy 1225 -----CTCCACTGCCACGGTG 1240
Db 600 helleSerLysThrCysAsnPheIleLeuIleThrThrCysAsnLeuPheCysSerThr 620
Qy 1241 AGGACACTTGGTCCCTGAC----- 1260
Db 620 hrLysThrAlaValProAspIleTyrArgAlaPheValMetAsnAspGluThrArgGlnM 640
Qy 1261 -----GGGAGCCAGCTACTTTCGCATAAGACCTGGGAGCTACAGATGCTATGTCCACAG 1315
Db 640 etSerGlyThrProCysGluAsnSerArgSerPheLeuSerSerAsnThrLeuSerSer- 659
Qy 1316 GAGCCCCACGAATCCTGTCCCGCTGGAGGCTGTGAGAAGGATGCTGGGGATAAGCCCT 1375
Db 660 --SerProGluAsn-----AlaProGluIleProL 669
Qy 1376 TAGCACCCAGTTAGACACCTCCAAAGAACACAGGCCCTGATGCAAGATGGCAGATCTG 1435
Db 669 euLeuGluAsnLeuLysPheAsnHisAsnLysLeuTyrLeuPheHisIleLeuSerProA 689
Qy 1436 ATACCCATTAGAGCCCGAGAAATCTCTCTCT----- 1467
Db 689 snHisHis-----AsnAsnSerSerSerThrGlyGlyTyrAsnAsnGlyHisHisG 706
Qy 1468 --GGATCCCACTTTGCAGCAAAACCCACACCCAGCTCACACAGCAAAACAA 1518
Db 706 InGlyProGlyMetLeuProGluAspAlaProSerTyrValLysGluArgGln 723

RESULT 5
S59856
collagen alpha 1(III) chain precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 10-Apr-1996 #sequence.Revision 19-Apr-1996 #text_change 13-Aug-1999
C:Accession: S59856; S62120; S16373
R:Toman, P.D.; de Crombrughe, B.
Gene 147, 161-168, 1994
A:Title: The mouse type-III procollagen-encoding gene: genomic cloning and complete l
A:Reference number: S59856; MUID:95011609; PMID:7926795
A:Cross-references: EMBL:X52046
A:Molecule type: DNA
A:Residues: 1-1464 <TOM>
R:Toman, D.
submitted to the EMBL Data Library, November 1994
A:Reference number: S62120
A:Accession: S62120
A:Molecule type: DNA
A:Residues: 1-866, 'G', 868-1464 <TOA>
A:Cross-references: EMBL:X52046; NID:g575321; PIDN:CAA36279.1; PID:g575322
R:Metsaeranta, M.; Toman, D.; de Crombrughe, B.; Vuorio, E.
Biochim. Biophys. Acta 1089, 241-243, 1991
A:Title: Specific hybridization probes for mouse type I, II, III and IX collagen mRNA
A:Reference number: S16176; MUID:91274355; PMID:2054384
A:Accession: S16373
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1442-1464 <MET>
A:Cross-references: EMBL:X57983; NID:g50476; PIDN:CAA41048.1; PID:g50477
C:Genetics:
A:Introns: 29/1; 95/3; 112/3; 150/3; 175/3; 193/3; 211/3; 229/3; 247/3; 265/3; 283/3;
58/3; 673/3; 706/3; 742/3; 778/3; 796/3; 814/3; 850/3; 868/3; 886/3; 940/3; 97
```

C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: coiled coil; extracellular matrix
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-154/Domain: propeptide #status predicted <PRO>
 F:32-92/Domain: von Willebrand factor type C repeat homology <WVC>
 F:155-1464/Product: collagen alpha 1(III) chain #status predicted <MAT>
 F:1236-1464/Domain: fibrillar collagen carboxyl-terminal homology <FCC>

Alignment Scores:
 Pld No.: 1 77e-07 Length: 1464
 Score: 218.00 Matches: 160
 Percent Similarity: 31.77% Conservative: 37
 Best Local Similarity: 25.81% Mismatches: 220
 Query Match: 6.97% Indels: 203
 DB: 2 Gaps: 36

US-09-762-021a-1 (1-1710) x S59856 (1-1464)

```

QY 10 ACTGAAGACCCCTGCAGAA---GGCTCTGGAGGAAGCTGGAGCAAAAGACCTCGACT 66
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 645 ThrGlyGlyProProGlyGluAsnGlyLysProGlyGluProGlyProLysGlyGluVal 664
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 67 TGGAGCCCTTCAGCCAGCCAGAGATGGAGGGGCTGTATGGAAGGCCCTCC 126
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 665 GlyAlaProGlyAlaProGlyGlyLysGlyAspSerGlyAlaProGlyGluArgGlyPro 684
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 127 TATGGAGCA-----GGCAGCGTATCTGGAGCGGGGATCCCTCCAGAACAGCC 174
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 685 ProGlyThrAlaGlyIleProGlyAlaArgGlyGlyAlaGlyProGlyProGly----- 701
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 175 CCACGAGGACCCCTAGA-----GCACAGCTCCACCATCCCAAGGCCCTCCAG 228
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 702 ---ProGluGlyGlyLysGlyProAlaGlyProGlyProGlyProGlyAlaSerGlySer 720
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 229 CCACACCACTGCCCGAGAACCAAGTGCCTTTACTTGCCTCTCCAGCGGCTCTCTTC 288
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 721 Pro-----GlyLeuGlnGly 725
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 289 CCCCAGGAGCCAGAGGAGGAGGAGTGTGTAACCATGCTTAAGGGACATTTGAGCT 348
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 726 MetProGlyGluArgGlyGlyProGlySer-----ProGlyProLysGlyGluLysGly 743
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 349 GTTCATGGGAAGCTGGA---GAAGGC---CCAGGCAAGACCCAGCAGGAAGAATTT 402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 744 GluProGlyGlyAlaGlyAlaAspGlyValProGlyLysAsp-----757
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 403 TGGGAAAAAANACAGGACCA---GGAGGCTCTACCCAGGACACAGTACATTTGCTGCTT 459
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 758 -----GlyProArgGlyProAlaGlyPro-----765
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 460 CCAGAAGATCAAGTACAGCTCTCAAGCTCTGGGAAGGCTGGCCACCTGGCTGAAGGAGAC 519
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 766 -----IleGlyProProGlyProAlaGlyGln-----ProGlyAsp 777
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 520 AAGTCCCTGAGCTGCTACACATCTCTTCAAGTCCCTGCACTTCATCTGCGCCAGGTG 579
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 778 LysGlyGluGlyGlySer-----ProGlyLeu---ProGlyIleAla 790
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 580 ---CCCTGAGGCTGGCTAGCAGCCCAAGTGTATCTCACCCCTCTCCACCCCTAAAGCTAT 636
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 791 GlyProArgGlyGlyProGlyGluArgGlyGluHisGlyProGlyProAlaGlyPhe 810
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 637 CAACCTGTACAGCTGCTGTCTAAGCCACCTCGAGATCAACCTTTGGATGGGGTTG----- 691
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 811 ProGlyAlaProGlyGlnAsn---GlyGluProGlyAlaLysGlyGluArgGlyAlaProGl 830
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 692 -----GGCCCCAGCTGGACCACTACTGCGGGCCGACTGGACAGCGCATGA 735
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 830 yGluLysGlyGluGlyGlyPro---ProGlyPro---AlaGlyProThrGlySerSerGlyP 849
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 736 GCCCTGCTCCACCAACCACTTCTCAGATGACTGGCAACTTCCAGAGCCCTCCAGCCA 795
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 849 roAlaGlyProProGlyProGlnGly---ValLysGlyGluArgGlySerPro----- 865
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

RESULT 6

CGH07L

collagen alpha 1(III) chain precursor - human

N:Alternate names: procollagen alpha 1(III) chain

```

QY 796 AGCACCCCTTAGGATACAGGACCCCTGTTTCCCTTCGGGGGAAGTCATAGCTTAGGGAG 855
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 866 ---GlyProProGlyThrAlaGly---PhePro---GlyGlyArg-----GlyL 879
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 856 CACCTCACACTTTCCTCAGGAGAGACACACAACCATGACCTCAGCCCTGGGAGCCCA- 914
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 879 euPro-----GlyProProG 884
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 915 -----ACTCAGGCCCTCCAGCCCAACCTGCCAGCCAGCCCTCAAAATGCA 963
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 884 lyAsnAsnGlyAsnProGlyProGlyProSer-----895
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 964 AGTCTGTACGAGTTTGAAGCTAGAACCCACGCGCAACTGACTGTGTCACAGGAGAGAA 1023
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db - 896 -----GlyAla-ProGlyLysAsp---GlyProProGlyPro 906
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1024 GCTGAGGTTTGGGA---CCACAGCAAGGGTGTGGTGTGAAGATGAGCGGGAGC 1080
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 907 AlaGlyAsnSerGlySerProGlyAsnProGlyIleAlaGlyProLys---GlyAspAla 925
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1081 GAGCGCTACATTCCAAGCAACATCTCGAGCCCTACAGCCGCGGAGCCCTGGGACCCA 1140
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 926 GlyGlnProGlyGluLysGlyProProGlyAla-----GlnGlyProProGlySerPro 943
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1141 GGGCAGCTCACCTCTCGGGTTTCCAACTGCTTCGACTTAGCTC-----1182
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 944 GlyProLeuGlyIleAlaGlyLeuThrGlyAlaAlaGlyLeuAlaGlyProGlyMet 963
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1183 -----GAGCCCTGAAGAGTTCACAGCTCGCTGCAGGACAGAACATTT 1224
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 964 ProGlyProArgGlySerProGlyProGlnGlyIleLysGlyGluSerGlyLysProGly 983
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1225 CTCCTACTGCCAGGCTGAGGACACTTGGGTCCCT-----1257
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 984 AlaSerGlyHisAsnGlyGluArgGlyProProGlyProGlnGlyLeuProGlyGlnPro 1003
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1258 -----GACGGGAGCCAGCTACT 1275
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1004 GlyThrAlaGlyGluProGlyArgAspGlyAsnProGlySerAspGlyGlnProGlyArg 1023
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1276 TCGCATAGACCTGGGAGCTACAGATGCTATGTCACAGGA-----1317
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1024 AspGlySerProGlyLysGlyAspArgGlyGlyLysGlySerProGlyAlaProGly 1043
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1318 -----GGCCCCAGCAATCTGTCTCCCGCTGGAGGCTGTGAGAGGATGCT 1362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1044 AlaProGlyHisProGlyProProGlyProValGlyProSerGlyLysSerGlyAspArg 1063
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1363 GGGGATAAG-----CCCTTAGGCACCACTTTAGACACCTTCCAAAGAACCCAGGCC 1410
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1064 GlyGluThrGlyProAlaGlyProSerGlyAlaProGlyProAlaGlyAlaArgGlyAla 1083
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1411 CGCTGTATGCAAGATGGCAGATCTGATACCCATTTAGAGCCCGCAGAAATCTCTTCTGGA 1470
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1084 ProGlyProGlnGlyProArgGlyAspLysGlyGluThrGlyGluArgGlySerAsnGly 1103
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1471 TCCCAAG-----TTTGCAGCAAAACCCACACCCACCTCACACAGCAAAACAA 1518
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1104 IleLysGlyHisArgGlyPheProGlyAsnProGlyPro-----1116
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1519 TGGACAGCCCGCAGAGCTGAAGCAACAGATGTCCCTTCTGGCTGTCTTGGAGCCCTCCCA 1578
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1117 -----ProGlySerProGlyAlaAlaGlyHisGlnGlyAlaIleGlySer----- 1131
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 1579 GTAACCACTATTATTATTTACCTCTTCCCAAACTTGGAGCATTTTGCCTAGG 1632
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1132 -----ProGlyProAlaGlyProArg 1138
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

C:Species: Homo sapiens (man)
 C:Date: 24-Apr-1984 #sequence_revision 01-Sep-1995 #text_change 21-Jul-2000
 C:Accession: S05272; S04642; PE0011; S01726; S04887; A90399; A94562; I51868; S59511; A90399; D.J.
 R:Prockop, D.J.
 Submitted to the EMBL Data Library, February 1989
 A:Reference number: S05272
 A:Accession: S05272
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-1240, 'V', 1242-1466 <PRC>
 A:Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
 R:Ala-Kokko, L.; Kontusaari, S.; Baldwin, C.T.; Kuivaniemi, H.; Prockop, D.J.
 Biochem. J. 260, 509-516, 1989
 A:Title: Structure of cDNA clones coding for the entire prepro-alpha(III) chain of human procollagen.
 A:Reference number: S04642; MUID:89350838; PMID:2764886
 A:Accession: S04642
 A:Molecule type: mRNA
 A:Residues: 1-1196 <ALA>
 A:Cross-references: EMBL:X14420; NID:g30057; PIDN:CAA32583.1; PID:g30058
 A:Note: the complete sequence is not shown
 R:Benson-Chanda, V.; Su, M.W.; Weil, D.; Chu, M.L.; Ramirez, F.
 Gene 78, 255-265, 1989
 A:Title: Cloning and analysis of the 5' portion of the human type-III procollagen gene (COL3A1).
 A:Reference number: PE0011; MUID:89378752; PMID:2777083
 A:Accession: PE0011
 A:Molecule type: DNA
 A:Residues: 1-176 <BEN>
 A:Cross-references: GB:M26939; NID:g180813; PIDN:AAA52040.1; PID:g180814
 R:Toman, P.D.; Ricca, G.A.; de Crombrughe, B.
 Nucleic Acids Res. 16, 7201, 1988
 A:Title: Nucleotide sequence of a cDNA coding for the amino-terminal region of human procollagen type III.
 A:Reference number: S01726; MUID:88303360; PMID:3405773
 A:Accession: S01726
 A:Molecule type: mRNA
 A:Residues: 1-170 <TOM>
 A:Cross-references: EMBL:X07240; NID:g30060; PIDN:CAA30229.1; PID:g30061
 A:Note: the authors translated the codon CAG for residue 154 as His
 R:Janeczko, R.A.; Ramirez, F.
 Nucleic Acids Res. 17, 6742, 1989
 A:Title: Nucleotide and amino acid sequences of the entire human alpha-1 (III) collagen.
 A:Reference number: S04887; MUID:89386015; PMID:2780304
 A:Accession: S04887
 A:Molecule type: mRNA
 A:Residues: 149-163, 'G', 164-240, 'D', 242-471, 'D', 473-487, 'L', 489, 'S', 491-613, 'Y', 615-634, 'Y', 635-636, 'Y', 637-638, 'Y', 639-640, 'Y', 641-642, 'Y', 643-644, 'Y', 645-646, 'Y', 647-648, 'Y', 649-650, 'Y', 651-652, 'Y', 653-654, 'Y', 655-656, 'Y', 657-658, 'Y', 659-660, 'Y', 661-662, 'Y', 663-664, 'Y', 665-666, 'Y', 667-668, 'Y', 669-670, 'Y', 671-672, 'Y', 673-674, 'Y', 675-676, 'Y', 677-678, 'Y', 679-680, 'Y', 681-682, 'Y', 683-684, 'Y', 685-686, 'Y', 687-688, 'Y', 689-690, 'Y', 691-692, 'Y', 693-694, 'Y', 695-696, 'Y', 697-698, 'Y', 699-700, 'Y', 701-702, 'Y', 703-704, 'Y', 705-706, 'Y', 707-708, 'Y', 709-710, 'Y', 711-712, 'Y', 713-714, 'Y', 715-716, 'Y', 717-718, 'Y', 719-720, 'Y', 721-722, 'Y', 723-724, 'Y', 725-726, 'Y', 727-728, 'Y', 729-730, 'Y', 731-732, 'Y', 733-734, 'Y', 735-736, 'Y', 737-738, 'Y', 739-740, 'Y', 741-742, 'Y', 743-744, 'Y', 745-746, 'Y', 747-748, 'Y', 749-750, 'Y', 751-752, 'Y', 753-754, 'Y', 755-756, 'Y', 757-758, 'Y', 759-760, 'Y', 761-762, 'Y', 763-764, 'Y', 765-766, 'Y', 767-768, 'Y', 769-770, 'Y', 771-772, 'Y', 773-774, 'Y', 775-776, 'Y', 777-778, 'Y', 779-780, 'Y', 781-782, 'Y', 783-784, 'Y', 785-786, 'Y', 787-788, 'Y', 789-790, 'Y', 791-792, 'Y', 793-794, 'Y', 795-796, 'Y', 797-798, 'Y', 799-800, 'Y', 801-802, 'Y', 803-804, 'Y', 805-806, 'Y', 807-808, 'Y', 809-810, 'Y', 811-812, 'Y', 813-814, 'Y', 815-816, 'Y', 817-818, 'Y', 819-820, 'Y', 821-822, 'Y', 823-824, 'Y', 825-826, 'Y', 827-828, 'Y', 829-830, 'Y', 831-832, 'Y', 833-834, 'Y', 835-836, 'Y', 837-838, 'Y', 839-840, 'Y', 841-842, 'Y', 843-844, 'Y', 845-846, 'Y', 847-848, 'Y', 849-850, 'Y', 851-852, 'Y', 853-854, 'Y', 855-856, 'Y', 857-858, 'Y', 859-860, 'Y', 861-862, 'Y', 863-864, 'Y', 865-866, 'Y', 867-868, 'Y', 869-870, 'Y', 871-872, 'Y', 873-874, 'Y', 875-876, 'Y', 877-878, 'Y', 879-880, 'Y', 881-882, 'Y', 883-884, 'Y', 885-886, 'Y', 887-888, 'Y', 889-890, 'Y', 891-892, 'Y', 893-894, 'Y', 895-896, 'Y', 897-898, 'Y', 899-900, 'Y', 901-902, 'Y', 903-904, 'Y', 905-906, 'Y', 907-908, 'Y', 909-910, 'Y', 911-912, 'Y', 913-914, 'Y', 915-916, 'Y', 917-918, 'Y', 919-920, 'Y', 921-922, 'Y', 923-924, 'Y', 925-926, 'Y', 927-928, 'Y', 929-930, 'Y', 931-932, 'Y', 933-934, 'Y', 935-936, 'Y', 937-938, 'Y', 939-940, 'Y', 941-942, 'Y', 943-944, 'Y', 945-946, 'Y', 947-948, 'Y', 949-950, 'Y', 951-952, 'Y', 953-954, 'Y', 955-956, 'Y', 957-958, 'Y', 959-960, 'Y', 961-962, 'Y', 963-964, 'Y', 965-966, 'Y', 967-968, 'Y', 969-970, 'Y', 971-972, 'Y', 973-974, 'Y', 975-976, 'Y', 977-978, 'Y', 979-980, 'Y', 981-982, 'Y', 983-984, 'Y', 985-986, 'Y', 987-988, 'Y', 989-990, 'Y', 991-992, 'Y', 993-994, 'Y', 995-996, 'Y', 997-998, 'Y', 999-1000, 'Y', 1001-1002, 'Y', 1003-1004, 'Y', 1005-1006, 'Y', 1007-1008, 'Y', 1009-1010, 'Y', 1011-1012, 'Y', 1013-1014, 'Y', 1015-1016, 'Y', 1017-1018, 'Y', 1019-1020, 'Y', 1021-1022, 'Y', 1023-1024, 'Y', 1025-1026, 'Y', 1027-1028, 'Y', 1029-1030, 'Y', 1031-1032, 'Y', 1033-1034, 'Y', 1035-1036, 'Y', 1037-1038, 'Y', 1039-1040, 'Y', 1041-1042, 'Y', 1043-1044, 'Y', 1045-1046, 'Y', 1047-1048, 'Y', 1049-1050, 'Y', 1051-1052, 'Y', 1053-1054, 'Y', 1055-1056, 'Y', 1057-1058, 'Y', 1059-1060, 'Y', 1061-1062, 'Y', 1063-1064, 'Y', 1065-1066, 'Y', 1067-1068, 'Y', 1069-1070, 'Y', 1071-1072, 'Y', 1073-1074, 'Y', 1075-1076, 'Y', 1077-1078, 'Y', 1079-1080, 'Y', 1081-1082, 'Y', 1083-1084, 'Y', 1085-1086, 'Y', 1087-1088, 'Y', 1089-1090, 'Y', 1091-1092, 'Y', 1093-1094, 'Y', 1095-1096, 'Y', 1097-1098, 'Y', 1099-1100, 'Y', 1101-1102, 'Y', 1103-1104, 'Y', 1105-1106, 'Y', 1107-1108, 'Y', 1109-1110, 'Y', 1111-1112, 'Y', 1113-1114, 'Y', 1115-1116, 'Y', 1117-1118, 'Y', 1119-1120, 'Y', 1121-1122, 'Y', 1123-1124, 'Y', 1125-1126, 'Y', 1127-1128, 'Y', 1129-1130, 'Y', 1131-1132, 'Y', 1133-1134, 'Y', 1135-1136, 'Y', 1137-1138, 'Y', 1139-1140, 'Y', 1141-1142, 'Y', 1143-1144, 'Y', 1145-1146, 'Y', 1147-1148, 'Y', 1149-1150, 'Y', 1151-1152, 'Y', 1153-1154, 'Y', 1155-1156, 'Y', 1157-1158, 'Y', 1159-1160, 'Y', 1161-1162, 'Y', 1163-1164, 'Y', 1165-1166, 'Y', 1167-1168, 'Y', 1169-1170, 'Y', 1171-1172, 'Y', 1173-1174, 'Y', 1175-1176, 'Y', 1177-1178, 'Y', 1179-1180, 'Y', 1181-1182, 'Y', 1183-1184, 'Y', 1185-1186, 'Y', 1187-1188, 'Y', 1189-1190, 'Y', 1191-1192, 'Y', 1193-1194, 'Y', 1195-1196, 'Y', 1197-1198, 'Y', 1199-1200, 'Y', 1201-1202, 'Y', 1203-1204, 'Y', 1205-1206, 'Y', 1207-1208, 'Y', 1209-1210, 'Y', 1211-1212, 'Y', 1213-1214, 'Y', 1215-1216, 'Y', 1217-1218, 'Y', 1219-1220, 'Y', 1221-1222, 'Y', 1223-1224, 'Y', 1225-1226, 'Y', 1227-1228, 'Y', 1229-1230, 'Y', 1231-1232, 'Y', 1233-1234, 'Y', 1235-1236, 'Y', 1237-1238, 'Y', 1239-1240, 'Y', 1241-1242, 'Y', 1243-1244, 'Y', 1245-1246, 'Y', 1247-1248, 'Y', 1249-1250, 'Y', 1251-1252, 'Y', 1253-1254, 'Y', 1255-1256, 'Y', 1257-1258, 'Y', 1259-1260, 'Y', 1261-1262, 'Y', 1263-1264, 'Y', 1265-1266, 'Y', 1267-1268, 'Y', 1269-1270, 'Y', 1271-1272, 'Y', 1273-1274, 'Y', 1275-1276, 'Y', 1277-1278, 'Y', 1279-1280, 'Y', 1281-1282, 'Y', 1283-1284, 'Y', 1285-1286, 'Y', 1287-1288, 'Y', 1289-1290, 'Y', 1291-1292, 'Y', 1293-1294, 'Y', 1295-1296, 'Y', 1297-1298, 'Y', 1299-1300, 'Y', 1301-1302, 'Y', 1303-1304, 'Y', 1305-1306, 'Y', 1307-1308, 'Y', 1309-1310, 'Y', 1311-1312, 'Y', 1313-1314, 'Y', 1315-1316, 'Y', 1317-1318, 'Y', 1319-1320, 'Y', 1321-1322, 'Y', 1323-1324, 'Y', 1325-1326, 'Y', 1327-1328, 'Y', 1329-1330, 'Y', 1331-1332, 'Y', 1333-1334, 'Y', 1335-1336, 'Y', 1337-1338, 'Y', 1339-1340, 'Y', 1341-1342, 'Y', 1343-1344, 'Y', 1345-1346, 'Y', 1347-1348, 'Y', 1349-1350, 'Y', 1351-1352, 'Y', 1353-1354, 'Y', 1355-1356, 'Y', 1357-1358, 'Y', 1359-1360, 'Y', 1361-1362, 'Y', 1363-1364, 'Y', 1365-1366, 'Y', 1367-1368, 'Y', 1369-1370, 'Y', 1371-1372, 'Y', 1373-1374, 'Y', 1375-1376, 'Y', 1377-1378, 'Y', 1379-1380, 'Y', 1381-1382, 'Y', 1383-1384, 'Y', 1385-1386, 'Y', 1387-1388, 'Y', 1389-1390, 'Y', 1391-1392, 'Y', 1393-1394, 'Y', 1395-1396, 'Y', 1397-1398, 'Y', 1399-1400, 'Y', 1401-1402, 'Y', 1403-1404, 'Y', 1405-1406, 'Y', 1407-1408, 'Y', 1409-1410, 'Y', 1411-1412, 'Y', 1413-1414, 'Y', 1415-1416, 'Y', 1417-1418, 'Y', 1419-1420, 'Y', 1421-1422, 'Y', 1423-1424, 'Y', 1425-1426, 'Y', 1427-1428, 'Y', 1429-1430, 'Y', 1431-1432, 'Y', 1433-1434, 'Y', 1435-1436, 'Y', 1437-1438, 'Y', 1439-1440, 'Y', 1441-1442, 'Y', 1443-1444, 'Y', 1445-1446, 'Y', 1447-1448, 'Y', 1449-1450, 'Y', 1451-1452, 'Y', 1453-1454, 'Y', 1455-1456, 'Y', 1457-1458, 'Y', 1459-1460, 'Y', 1461-1462, 'Y', 1463-1464, 'Y', 1465-1466, 'Y', 1467-1468, 'Y', 1469-1470, 'Y', 1471-1472, 'Y', 1473-1474, 'Y', 1475-1476, 'Y', 1477-1478, 'Y', 1479-1480, 'Y', 1481-1482, 'Y', 1483-1484, 'Y', 1485-1486, 'Y', 1487-1488, 'Y', 1489-1490, 'Y', 1491-1492, 'Y', 1493-1494, 'Y', 1495-1496, 'Y', 1497-1498, 'Y', 1499-1500, 'Y', 1501-1502, 'Y', 1503-1504, 'Y', 1505-1506, 'Y', 1507-1508, 'Y', 1509-1510, 'Y', 1511-1512, 'Y', 1513-1514, 'Y', 1515-1516, 'Y', 1517-1518, 'Y', 1519-1520, 'Y', 1521-1522, 'Y', 1523-1524, 'Y', 1525-1526, 'Y', 1527-1528, 'Y', 1529-1530, 'Y', 1531-1532, 'Y', 1533-1534, 'Y', 1535-1536, 'Y', 1537-1538, 'Y', 1539-1540, 'Y', 1541-1542, 'Y', 1543-1544, 'Y', 1545-1546, 'Y', 1547-1548, 'Y', 1549-1550, 'Y', 1551-1552, 'Y', 1553-1554, 'Y', 1555-1556, 'Y', 1557-1558, 'Y', 1559-1560, 'Y', 1561-1562, 'Y', 1563-1564, 'Y', 1565-1566, 'Y', 1567-1568, 'Y', 1569-1570, 'Y', 1571-1572, 'Y', 1573-1574, 'Y', 1575-1576, 'Y', 1577-1578, 'Y', 1579-1580, 'Y', 1581-1582, 'Y', 1583-1584, 'Y', 1585-1586, 'Y', 1587-1588, 'Y', 1589-1590, 'Y', 1591-1592, 'Y', 1593-1594, 'Y', 1595-1596, 'Y', 1597-1598, 'Y', 1599-1600, 'Y', 1601-1602, 'Y', 1603-1604, 'Y', 1605-1606, 'Y', 1607-1608, 'Y', 1609-1610, 'Y', 1611-1612, 'Y', 1613-1614, 'Y', 1615-1616, 'Y', 1617-1618, 'Y', 1619-1620, 'Y', 1621-1622, 'Y', 1623-1624, 'Y', 1625-1626, 'Y', 1627-1628, 'Y', 1629-1630, 'Y', 1631-1632, 'Y', 1633-1634, 'Y', 1635-1636, 'Y', 1637-1638, 'Y', 1639-1640, 'Y', 1641-1642, 'Y', 1643-1644, 'Y', 1645-1646, 'Y', 1647-1648, 'Y', 1649-1650, 'Y', 1651-1652, 'Y', 1653-1654, 'Y', 1655-1656, 'Y', 1657-1658, 'Y', 1659-1660, 'Y', 1661-1662, 'Y', 1663-1664, 'Y', 1665-1666, 'Y', 1667-1668, 'Y', 1669-1670, 'Y', 1671-1672, 'Y', 1673-1674, 'Y', 1675-1676, 'Y', 1677-1678, 'Y', 1679-1680, 'Y', 1681-1682, 'Y', 1683-1684, 'Y', 1685-1686, 'Y', 1687-1688, 'Y', 1689-1690, 'Y', 1691-1692, 'Y', 1693-1694, 'Y', 1695-1696, 'Y', 1697-1698, 'Y', 1699-1700, 'Y', 1701-1702, 'Y', 1703-1704, 'Y', 1705-1706, 'Y', 1707-1708, 'Y', 1709-1710, 'Y', 1711-1712, 'Y', 1713-1714, 'Y', 1715-1716, 'Y', 1717-1718, 'Y', 1719-1720, 'Y', 1721-1722, 'Y', 1723-1724, 'Y', 1725-1726, 'Y', 1727-1728, 'Y', 1729-1730, 'Y', 1731-1732, 'Y', 1733-1734, 'Y', 1735-1736, 'Y', 1737-1738, 'Y', 1739-1740, 'Y', 1741-1742, 'Y', 1743-1744, 'Y', 1745-1746, 'Y', 1747-1748, 'Y', 1749-1750, 'Y', 1751-1752, 'Y', 1753-1754, 'Y', 1755-1756, 'Y', 1757-1758, 'Y', 1759-1760, 'Y', 1761-1762, 'Y', 1763-1764, 'Y', 1765-1766, 'Y', 1767-1768, 'Y', 1769-1770, 'Y', 1771-1772, 'Y', 1773-1774, 'Y', 1775-1776, 'Y', 1777-1778, 'Y', 1779-1780, 'Y', 1781-1782, 'Y', 1783-1784, 'Y', 1785-1786, 'Y', 1787-1788, 'Y', 1789-1790, 'Y', 1791-1792, 'Y', 1793-1794, 'Y', 1795-1796, 'Y', 1797-1798, 'Y', 1799-1800, 'Y', 1801-1802, 'Y', 1803-1804, 'Y', 1805-1806, 'Y', 1807-1808, 'Y', 1809-1810, 'Y', 1811-1812, 'Y', 1813-1814, 'Y', 1815-1816, 'Y', 1817-1818, 'Y', 1819-1820, 'Y', 1821-1822, 'Y', 1823-1824, 'Y', 1825-1826, 'Y', 1827-1828, 'Y', 1829-1830, 'Y', 1831-1832, 'Y', 1833-1834, 'Y', 1835-1836, 'Y', 1837-1838, 'Y', 1839-1840, 'Y', 1841-1842, 'Y', 1843-1844, 'Y', 1845-1846, 'Y', 1847-1848, 'Y', 1849-1850, 'Y', 1851-1852, 'Y', 1853-1854, 'Y', 1855-1856, 'Y', 1857-1858, 'Y', 1859-1860, 'Y', 1861-1862, 'Y', 1863-1864, 'Y', 1865-1866, 'Y', 1867-1868, 'Y', 1869-1870, 'Y', 1871-1872, 'Y', 1873-1874, 'Y', 1875-1876, 'Y', 1877-1878, 'Y', 1879-1880, 'Y', 1881-1882, 'Y', 1883-1884, 'Y', 1885-1886, 'Y', 1887-1888, 'Y', 1889-1890, 'Y', 1891-1892, 'Y', 1893-1894, 'Y', 1895-1896, 'Y', 1897-1898, 'Y', 1899-1900, 'Y', 1901-1902, 'Y', 1903-1904, 'Y', 1905-1906, 'Y', 1907-1908, 'Y', 1909-1910, 'Y', 1911-1912, 'Y', 1913-1914, 'Y', 1915-1916, 'Y', 1917-1918, 'Y', 1919-1920, 'Y', 1921-1922, 'Y', 1923-1924, 'Y', 1925-1926, 'Y', 1927-1928, 'Y', 1929-1930, 'Y', 1931-1932, 'Y', 1933-1934, 'Y', 1935-1936, 'Y', 1937-1938, 'Y', 1939-1940, 'Y', 1941-1942, 'Y', 1943-1944, 'Y', 1945-1946, 'Y', 1947-1948, 'Y', 1949-1950, 'Y', 1951-1952, 'Y', 1953-1954, 'Y', 1955-1956, 'Y', 1957-1958, 'Y', 1959-1960, 'Y', 1961-1962, 'Y', 1963-1964, 'Y', 1965-1966, 'Y', 1967-1968, 'Y', 1969-1970, 'Y', 1971-1972, 'Y', 1973-1974, 'Y', 1975-1976, 'Y', 1977-1978, 'Y', 1979-1980, 'Y', 1981-1982, 'Y', 1983-1984, 'Y', 1985-1986, 'Y', 1987-1988, 'Y', 1989-1990, 'Y', 1991-1992, 'Y', 1993-1994, 'Y', 1995-1996, 'Y', 1997-1998, 'Y', 1999-2000, 'Y', 2001-2002, 'Y', 2003-2004, 'Y', 2005-2006, 'Y', 2007-2008, 'Y', 2009-2010, 'Y', 2011-2012, 'Y', 2013-2014, 'Y', 2015-2016, 'Y', 2017-2018, 'Y', 2019-2020, 'Y', 2021-2022, 'Y', 2023-2024, 'Y', 2025-2026, 'Y', 2027-2028, 'Y', 2029-2030, 'Y', 2031-2032, 'Y', 2033-2034, 'Y', 2035-2036, 'Y', 2037-2038, 'Y', 2039-2040, 'Y', 2041-2042, 'Y', 2043-2044, 'Y', 2045-2046, 'Y', 2047-2048, 'Y', 2049-2050, 'Y', 2051-2052, 'Y', 2053-2054, 'Y', 2055-2056, 'Y', 2057-2058, 'Y', 2059-2060, 'Y', 2061-2062, 'Y', 2063-2064, 'Y', 2065-2066, 'Y', 2067-2068, 'Y', 2069-2070, 'Y', 2071-2072, 'Y', 2073-2074, 'Y', 2075-2076, 'Y', 2077-2078, 'Y', 2079-2080, 'Y', 2081-2082, 'Y', 2083-2084, 'Y', 2085-2086, 'Y', 2087-2088, 'Y', 2089-2090, 'Y', 2091-2092, 'Y', 2093-2094, 'Y', 2095-2096, 'Y', 2097-2098, 'Y', 2099-2100, 'Y', 2101-2102, 'Y', 2103-2104, 'Y', 2105-2106, 'Y', 2107-2108, 'Y', 2109-2110, 'Y', 2111-2112, 'Y', 2113-2114, 'Y', 2115-2116, 'Y', 2117-2118, 'Y', 2119-2120, 'Y', 2121-2122, 'Y', 2123-2124, 'Y', 2125-2126, 'Y', 2127-2128, 'Y', 2129-2130, 'Y', 2131-2132, 'Y', 2133-2134, 'Y', 2135-2136, 'Y', 2137-2138, 'Y', 2139-2140, 'Y', 2141-2142, 'Y', 2143-2144, 'Y', 2145-2146, 'Y', 2147-2148, 'Y', 2149-2150, 'Y', 2151-2152, 'Y', 2153-2154, 'Y', 2155-2156, 'Y', 2157-2158, 'Y', 2159-2160, 'Y', 2161-2162, 'Y', 2163-2164, 'Y', 2165-2166, 'Y', 2167-2168, 'Y', 2169-2170, 'Y', 2171-2172, 'Y', 2173-2174, 'Y', 2175-2176, 'Y', 2177-2178, 'Y', 2179-2180, 'Y', 2181-2182, 'Y', 2183-2184, 'Y', 2185-2186, 'Y', 2187-2188, 'Y', 2189-2190, 'Y', 2191-2192, 'Y', 2193-2194, 'Y', 2195-2196, 'Y', 2197-2198, 'Y', 2199-2200, 'Y', 2201-2202, 'Y', 2203-2204, 'Y', 2205-2206, 'Y', 2207-2208, 'Y', 2209-2210, 'Y', 2211-2212, 'Y', 2213-2214, 'Y', 2215-2216, 'Y', 2217-2218, 'Y', 2219-2220, 'Y', 2221-2222, 'Y', 2223-2224, 'Y', 2225-2226, 'Y', 2227-2228, 'Y', 2229-2230, 'Y', 2231-2232, 'Y', 2233-2234, 'Y', 2235-2236, 'Y', 2237-2238, 'Y', 2239-2240, 'Y', 2241-2242, 'Y', 2243-2244, 'Y', 2245-2246, 'Y', 2247-2248, 'Y', 2249-2250, 'Y', 2251-2252, 'Y', 2253-2254, 'Y', 2255-2256, 'Y', 2257-2258, 'Y', 2259

A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1161-1200 <MIS>
 A:Cross-references: GB:M13146; NID:g180415; PIDN:AAA52003.1; PID:g180416
 R:Emanuel, B.S.; Cannizzaro, L.A.; Sever, J.M.; Myers, J.C.
 Proc. Natl. Acad. Sci. U.S.A. 82, 3385-3389, 1985
 A:Title: Human alpha 1(III) and alpha 2(V) procollagen genes are located on the long arm
 A:Reference number: 159025; MUID:85216505; PMID:3858826
 A:Accession: I79359
 A>Status: translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1165-1196 <MA>
 A:Cross-references: GB:M11134; NID:g180417; PIDN:AAA52004.1; PID:g180418
 R:Chu, M.L.; Weil, D.; de Wet, W.; Bernard, M.; Sippola, M.; Ramirez, F.
 J. Biol. Chem. 260, 4357-4363, 1985
 A:Title: Isolation of cDNA and genomic clones encoding human pro-alpha(III) collagen. B
 A:Reference number: A92516; MUID:85157600; PMID:2579949
 A:Accession: A92516
 A:Molecule type: DNA
 A:Residues: 1176-1240, 'V', 1242-1356, 'P', 1358-1466 <CHU>
 A:Cross-references: GB:M10615; GB:M10793; GB:M10794; GB:M10795; GB:M10796; GB:M10797; GB
 A:Experimental source: liver
 A>Note: the authors translated the codon TTC for residue 1057 as Tyr; the codons given f
 ation
 C:Comment: Prolines and lysines at the third position of the tripeptide repeating unit
 C:Genetics:
 A:Gene: GDB:COL3A1
 A:Cross-references: GDB:118729; OMIM:120180
 A:Map position: 2q31-2q31
 A:Introns: 27/1; 94/3; 111/3; 149/3; 176/3; 554/3; 587/3; 1175/3; 1275/1; 1337/3; 1418/3
 A>Note: the list of introns is incomplete; defects in this gene can result in Ehlers-Dan
 C:Complex: type III collagen is a homotrimer of monomers initially linked by disulfide b
 er of their length, is formed with desmosine cross-links made from lysine and allysine r
 C:Function:
 A:Description: structural component of extracellular fibrous polymer that maintains inte
 C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology;
 C:Keywords: coiled coil; Ehlers-Danlos syndrome; extracellular matrix; glycoprotein; hyd
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-153/Domain: amino-terminal propeptide #status predicted <PRO>
 F:31-91/Domain: von Willebrand factor type C repeat homology <VW>
 F:154-1221/Product: collagen alpha 1(III) chain #status predicted <MAT>
 F:154-167/Region: amino-terminal nonhelical telopeptide
 F:168-1196/Region: helical
 F:1091-1093/Region: cell attachment (R-G-D) motif
 F:1197-1221/Region: carboxyl-terminal nonhelical telopeptide
 F:1222-1466/Domain: carboxyl-terminal propeptide #status predicted <CPR>
 F:1238-1466/Domain: fibrillar collagen carboxyl-terminal homology <FCC>
 F:24/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicted
 F:153-154/Cleavage site: pro-Gln (procollagen N-endopeptidase) #status predicted
 F:154/Modified site: pyrrolidone carboxylic acid (Gln) (in mature form) #status predicte
 F:161,1212/Modified site: allysine (Lys) #status predicted
 F:263,284,860,977,1106/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:263/Binding site: carboxylate (Lys) (covalent) #status experimental
 F:584,1094/Modified site: 5-hydroxylysine (Lys) (partial) #status experimental
 F:948-949/Cleavage site: Gly-Ile (collagenase) #status experimental
 F:1106/Binding site: carboxylate (Lys) (covalent) #status predicted

Alignment Scores:

Pred. No.:	3,01e-06	Length:	1466
Score:	200.50	Matches:	157
Percent Similarity:	32.01%	Conservative:	37
Best Local Similarity:	25.91%	Mismatches:	224
Query Match:	6.41%	Indels:	188
DB:	1	Gaps:	34

US-09-762-021A-1 (1-1710) x CGH07L (1-1466)

QY	10	ACTGAAGACGACGCTGCAGAA---GGCTCTGGAGGAGAGCTGGAGCAAGACCTCGACT	66
Db	646	ThrGlyGlyProGlyGluAsnGlyLysProGlyGluProGlyGlyProGlyProGlyLysGlyAspAla	665
QY	67	TGGAGCCTTCACGCCAGGCCAGACAGATGGAGGGGGCCTGTCTATGAAAGCCGCTCCC	126

Db	666	GlyAlaProGlyAlaProGlyGlyLysGlyAspAlaGlyAlaProGlyGlyGluArgGlyPro	685
QY	127	TATGGA-----CGAGCGACGCTATCTGGAGCGGGGATCCCTCCAGAACGCC	174
Db	686	ProGlyLeuAlaGlyAlaProGlyLeuArgGlyGlyAlaGlyProGly-----	702
QY	175	CCACGAGAGACCCCTAGA-----GCACAGCCTCCACCATCCCAAGGCCCTCCGACG	228
Db	703	---ProGluGlyGlyLysGlyAlaAlaGlyProGlyProGlyAlaAlaGlyThr	721
QY	229	CCACACCACTGCCGCGAGAACCAAGTCCTTACTTCGCCTCTCTCAAGCGGTCCTCTTC	288
Db	722	Pro-----GlyLeuGlnGly	726
QY	289	CCCCGAGGACCCAGAGAGGAGGAGTGTGTAACCATGTCTTAAAGGACATTTGAGCT	348
Db	727	MetProGlyGluArgGlyGlyLeuGlySer-----ProGlyProLysGlyAsp-----	742
QY	349	GTTCATGGGAAAGCTGGAGAGGCCCGAGCAAGACAGCAGCAGGAGAAGAAATTTGGGA	408
Db	743	---LysGlyGluProGly---GlyProGlyAlaAspGlyVal-----Pro	755
QY	409	AAAAAACAAGGACCA---GGGAGGTCTCACCCAGGACACAGTACATTGACTGCTTCAGAA	465
Db	756	GlyLysAspGlyProArgGlyProThrGlyPro-----AlaProGlyLeu---ProGlyIleAlaGlyPro	766
QY	466	GATCAAGTACAGCTTCAACCTCTCTGGGAGGCTGGCCACCTGGCTGAAGGACACAAAGTCG	525
Db	767	-----IleGlyProProGlyProAlaGlyGln-----ProGlyAspLysGly	780
QY	526	CCCTGAGCTCGTACACATCTCTTCAAGTCCCTGAATCTTCTCTGCGCCAGGTG---CCC	582
Db	781	GluGlyGly-----AlaProGlyLeu---ProGlyIleAlaGlyPro	793
QY	583	TGAGCTGCGCTAGCAGGCCACAGTGTATCTACCCCTCTCTACCCCTAAAGCTATCAACCT	642
Db	794	ArgGlySerProGlyGluArgGlyGlyGluThrGlyProProGlyProAlaGlyPheProGly	813
QY	643	GCTACAGCTCTGCTTAAGCCACCTCGAGAGTAACCTTTGGATGGGGTTG-----	691
Db	814	AlaProGlyGlnAsn---GlyGluProGlyGlyLysGlyGlyGluArgGlyAlaProGlyGlu	833
QY	692	-----GGCCAGCCTGGACCACTAGCCGGCGCGCTGCACAGCGCATGAGCCCT	741
Db	833	sglyGluGlyGlyPro---ProGly---ValAlaGlyProProGlyGlySerGlyProAlaG	852
QY	742	GCCCTACCAACCCATCTCAGATGACTGGCAACTTCCAGAGCCCT-----	788
Db	852	lyProProGlyProGlnGly---ValLysGlyGluArgGlySerProGlyGlyProGlyA	871
QY	789	-----CCAGCAAGCACCCCTTAGGATACCA	813
Db	871	laAlaGlyPheProGlyAlaArgGlyLeuProGlyPro---ProGlySerAsnGlyAsnPro	890
QY	814	GGACCTGTTTCCCTTCGGCGGGGAGTCAAGTTAGGGAGCACCCTCACACTTTCCTCA	873
Db	891	GlyProProGlyProSerGly-----SerPro	899
QY	874	GGAGAGACACACACCATGACCTCA-----GCC	903
Db	900	GlyLysAspGlyProProGlyProAlaGlyAsnThrGlyAlaProGlySerProGlyVal	919
QY	904	TGGGGACCCCACTCCAGGCCCTCCAGCCC-----CAAAACC	939
Db	920	SerGlyProLysGlyAspAlaGlyGlnProGlyGlyLysGlySerProGlyAlaGlnGly	939
QY	940	TGCCCGACCGCCCTGAAATGCAAGTCTTTGACGAGTTTGAAGCTTAGGAACCCACGGA	999
Db	940	ProProGlyAlaProGlyProLeuGlyIleAlaGlyIleThrGlyAlaArgGlyLeuAla	959
QY	1000	ACTGACTGTGTCCAGGGGAGAGAGCTGTGGACCAAGCGGTGGTGGCT	1059


```
QY 930 GCCCAACCTGCC-----CAGCCAGCCCTGAAATGCAAGTCTTGTACGAGT 977
Db 638 erProProThrHisAsnThrAsnGlnProProMet----- 650
QY 978 TTGAAGCTAGGAACCCACGGGAACCTGACTGTGTCCAGGGAGAGAGCTGGAGGTTCTGG 1037
Db 651 --GlyAlaProThrGlnAlaProThrProSerSerGluThrThrGlnValProT 670
QY 1038 ACCACAGCAAGCGGTGGTGGTGTGAAGAATGAGCGGGAGCGGACGTACATTCCAA 1097
Db 670 hrProSer-----SerGluSerAspGln-----S 678
QY 1098 GCAACATCTGTGAGCCCTACAGCCGGGGAGCCCTGGGACCCAGTCACCTCTC 1157
Db 678 erGlnIleLeuSerProValGlnAlaProThrPro---ValGlnSerSerThrPro--- 695
QY 1158 GGGTTCCAATGCTTCGACTTACCTCGAGGCTGAAGAGTC-----ACAG 1202
Db 696 -----SerSerGluProThrGlnValProThrProSerSerSerG 709
QY 1203 ACTGGCTGCAGGACAGAACTTCTCCACTGCCACGGTGAGGACACTTGGTCCCTGACGG 1262
Db 709 luSerTyrGlnAlaProAsnLeuSerProValGlnAlaProThr----- 723
QY 1263 GGAGCCAGCTACTTCGCATAAGACCTGGGGAGCTACAGATGCTATGTCACAGAGGCC 1322
Db 724 -----ProValGlnAlaP 728
QY 1323 CAGCAATCTGTCCCGCTGGAGGCTGTACAGAGGATGCTGGGATAAGCCCTTAGGCAC 1382
Db 728 roThrThrSerSerGluThrSerGlnValProThrProSerSerGluSerAsnGlnSerP 748
QY 1383 CAGCTTTAGACACTCCCAAGAACCCAGGCCCGCTGTGTAAGATGGCAGATCTGATACCCA 1442
Db 748 roSerGlnAlaProThrProIle-----LeuGluProV 759
QY 1443 TTAGAGCCCGAGAAATCTCTTCCTGGATGCCAGTTCCAGCAAAACCCACCCAGCT 1502
Db 759 alHisAlaPro---ThrProAsnSerLysPro-----ValGlnSerProThrProSerS 776
QY 1503 CACACAGCAAAACATGACAGAGGCCAGGCTGACGCAACACATGTCCCTTCTGGCTG 1562
Db 776 erGluProValSerSerProGluGlnSerGlu-----G 787
QY 1563 TGTGTGAGCTCCAGTAACACCACTATTATTATTTACCTTTTCCCAAC 1612
Db 787 luValGluAlaProGluProThrProValAsnProSerSerValProSer 803

RESULT 8
QBEB3
BHLF1 protein - human herpesvirus 4 (strain B95-8)
C:Species: human herpesvirus 4, Epstein-Barr virus
C:Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 23-Aug-1997
C:Accession: A03742
R:Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A:Reference number: A93065; MUID:85035713; PMID:6092825
A:Accession: A03742
A:Molecule type: DNA
A:Residues: 1-660 <BAN>
R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
Nature 310, 207-211, 1984
A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A:Reference number: A03794; MUID:84270667; PMID:6087149
A:Contents: annotation; protein coding region
C:Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52
C:Superfamily: human herpesvirus 4 BHLF1 protein

Alignment Scores:
Pred. No.: 1,11e-05 Length: 660
Score: 192.00 Matches: 145
Percent Similarity: 32.28% Conservative: 49
```

```
Best Local Similarity: 24.13% Mismatches: 192
Query Match: 6.14% Indels: 215
DB: 1 Gaps: 36

US-09-762-021a-1 (1-1710) x QBEB3 (1-660)

QY 19 CAGCCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGACCTCGACTTGGAGGCTTCA 78
Db 177 ArgProSerGlyProThrGlyArgProAlaAlaProGlyAlaProGlyThrProAla 196
QY 79 GCCAGGCCAGACAGATGGAGGGCCCTGCTATGGAAGGCGCTCCCTATGTGACGAGC 138
Db 197 Ala-----TCTGGAGCCGGGATCCCTCCAGAACAGCC-----AlaValProSerGlyAlaThr 211
QY 139 ACGCTA-----TCTGGAGCCGGGATCCCTCCAGAACAGCC-----CACCA 180
Db 212 ProHisProGluArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProPro 231
QY 181 GAG-----GACCTAGA-----GCACAGCCCTCCACCATC 210
Db 232 GluArgGlnGluProArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGly 251
QY 193 -----GCACAGCCCTCCACCATC 210
Db 252 ProProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCys 271
QY 211 CCCAAGGCCCTGCCACGCCACACAGTCCCGAGAACCAAGTCTTACTCTGCCTCC 270
Db 272 ProArgSerAlaArgAsnPro---GlyCysProArgThrTrpArgArgSerGlyAla 290
QY 271 TCCAAG-----GCGTCTCTTTCCTCCCGAGGACCAGGAGGAGGAGCA 312
Db 291 GlnArgGlyHisProProProGlyAlaGlyGlnArgProSerGlyProThrGlyArg 310
QY 313 GGAAGT-----GCTGAACCATGCTCTTAAGGACATTTAGCT 348
Db 311 ProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAla 330
QY 349 GTT-----CATGGAAAGCTGGAGAGGCCAGGCAAGACAGCAG 390
Db 331 ValProSerGlyAlaThrProHisProGluArgGlySerGlyProAlaAspProProAla 350
QY 391 GAAGAAGAAATTTGGGAAAAAACAAGGACCCAGGAGGTCTCACCCAGGACAGATACAT 450
Db 351 AlaAlaArgLeuProProGluArgGlnGluPro-----CCTCTGGG 492
QY 451 TCAGTCTTCCAGAGATCAAGTACAGCTTCA-----CCTCTGGG 492
Db 362 ---ArgLeuProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProThr 380
QY 493 AAGGCTGCCACCTGGCTGAAGGAGACAAAGTCCCTCGAGCTCGTACACAT-----CCT 546
Db 381 ArgSerGlyAlaAlaAlaGln-----ArgThrHisArgArgPro 393
QY 547 CTTCAAGTCCCTGAA-----CTTCATCTCTGGCCAGGTGCCCTGAGCTGG 591
Db 394 ProGlyCysProArgSerAlaArgAsnProGly---CysProArgThrTrpArgArg 412
QY 592 -----CCTAGCAGCCCAAGTGTCTCACCCCTCTCCACCCCTCA 630
Db 413 SerGlyAlaGlnArgGlyHisProProProGlyAlaGlnArgProSerGlyProThr 432
QY 631 AGCTATCAACTGCTTACAGTCTGTCTTAAGCCACCTCAGAGTAACCTTTGGATGGGTT 690
Db 433 GlyGlyArgProAla---AlaPro----- 439
QY 691 GGGCCCACTTGA---CCACTAGCC---GGGCCGACTGGACAGCGCATGAGCCCTGCC 744
Db 440 --GlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAlaValProSerG 459
QY 745 CTACCAACCCACATTTCTCAGATGACTGGCACTTCCAGAGCCCTCCAGCAGCACCTT 804
Db 459 lAlaThrProHisProGlu---ArgGlySerGlyProAlaAspPro-ProAlaAlaAla 477
```

Qy 805 AGGATACCAGCACCTGTTTCCCTCGCGGGAAGTCATAGTTAGGGACACCTCACA 864
| | | | |
Db 478 ArgLeuPro-----ProGluArgGlnGluProArgLeuProGlnAspLeuAla 493
Qy 865 CTTTCCTCAGGAGACACACAACCATGACCTCAGCCTGGGACCCCAACTCCAGGCC 924
| | | | |
Db 494 AlaAlaGlnArgCysProAlaGlyProProProThrArgSerGlyAlaAlaAlaGlnArg 513
Qy 925 CTC-----CAGCCCCAAACCTGCCAGCCCTGAAATGCAAGTCTGTACGAGTT 978
| | | | |
Db 514 ThrHisArgArgProProGlyCysProArgSerAlaArgAsnProGlyCysProArgThr 533
Qy 979 TGAAGTAGGAACCCAGCGGAACCTGACTGTGCT-----CCAGGAGAGAAAGCTGGA 1029
| | | | |
Db 534 TrpArg---ArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlyGlnArg 552
Qy 1030 GTTCTGGACACACAAAGCGGTGGTGGCTGGTGAAGANTAGCGGGACGAGCGGCTA 1089
| | | | |
Db 553 ProSerGlyProThr-----GlyGlyArgProAlaAla 563
Qy 1090 CATTCGAACCAACATCTCTGGAGCCCTACAGCCGGGACCCCTGGGACCCAGGCCAGTC 1149
| | | | |
Db 564 -----ProGlyAlaPro-----GlyThrProAlaAlaProGlyPro----- 575
Qy 1150 ACCCTCTCGGGTTCCAATGCTTCGACTTAGCTCGAGGCGCTGAAGAGGTCACAGACTGGCT 1209
| | | | |
Db 576 -----GlyGlyGlyAlaAla----- 580
Qy 1210 CGAGCAGAGAACCTTCTCCACTGCCACGGTGAGGACACTTGGTCCCTGACGGGAGCCA 1269
| | | | |
Db 581 -----ValProSerGly----- 584
Qy 1270 GCTACTTCGCATAGACCTGGGAGCTACAGATGCTATGTCACAGGAGGCCCAACGAT 1329
| | | | |
Db 585 AlaThrProHisProGluArgGlySer-----GlyProAlaAsp 597
Qy 1330 CCTGTCCCGCTGGAGGCTGTCAAGAAGGATGCTGGGATAAGCCCTTAGGCACCACTTA 1389
| | | | |
Db 598 Pro-----ProAlaAlaAlaArgLeu 604
Qy 1390 GACACTCCAGAACACCAGGCCCTCGATGACAGATGCGCATGATACCCATTAGAGC 1449
| | | | |
Db 605 ProProGluArgGlnGluProArgLeu----- 613
Qy 1450 CCCGAGAAATCTCTCTGGATCCAGTTTCAGCAAAACCCACACCCACCTCACAG 1509
| | | | |
Db 614 ProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProThrArg----- 631
Qy 1510 CAAAAACAATGGCAGGCCCCAGAGCGTGAA---GCAAAACAGTGTCCCTTCTGGCTGT 1563
| | | | |
Db 632 -----SerGlyAlaAlaAlaGlnArgThrHisArgArgProProGlyCys 646
RESULT 9
T03166
probable immediate early protein - alcelaphine herpesvirus 1
C:Species: alcelaphine herpesvirus 1
C>Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 24-Nov-1999
C:Accession: T03166
R:Ensser, A.; Pflanz, R.; Fleckenstein, B.
J. Virol. 71, 6517-6525, 1997
A:Title: Primary structure of the alcelaphine herpesvirus 1 genome.
A:Reference number: 214840; MUID: 97404659; PMID: 9261371
A:Accession: T03166
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1300 <ENS>
A:Cross-references: EMBL:AF005370; NID:q2337967; PIDN:AAG58118.1; PID:q2338034
C:Superfamily: collagen alpha 2(I) chain; fibrillar collagen carboxyl-terminal homology
Alignment Scores:
Pred. No.: 1.39e-05 Length: 1300
Score: 191.00 Matches: 139

Percent Similarity: 29.89% Conservative: 26
Best Local Similarity: 25.18% Mismatches: 244
Query Match: 6.11% Indels: 144
DB: 2 Gaps: 29
US-09-762-021a-1 (1-1710) x T03166 (1-1300)
Qy 13 GAAGACCAGCCTGCAGAGGCTCTGGAGAAAGAGCTGGAGCAAGACACCTCGACTTGGAGG 72
| | | | |
Db 219 GluGlyProLysGlyGluGlyProGluGlyProGluGlyProGluGlyAspSer----- 236
Qy 73 CTTTCAGCAGCCAGGACACAGATGAGGGGCTCTATGAAAGGCCCTCCCTATGGA 132
| | | | |
Db 237 ProAspGly---ProGlyAlaGlnGluGlyPro---GluGlyLeuGluGlyProGluGly 254
Qy 133 CGAGCAGCCTATCTGGAGCCGGGATCCCTCCAGAACAGCCCA-----CCAGAGAC 186
| | | | |
Db 255 AspGluGlyProGluGlyProGluGlyProGluGlyGluGlyProGluGlyProGluGly 274
Qy 187 CCTAGACACAGCCTCCCAAGCCCTCCCAAGCCCTGCGCCACACACAGTCGCCGAGA 246
| | | | |
Db 275 ProLysGlyAspSerProAspGlyProGlyAlaGlnGluGlyProGluGly----- 291
Qy 247 ACCAAGTGCCTTTACTCTGCCTCTCCTCAAGCGGTCTCTTCCCGCGAGGACCCAGAGAG 306
| | | | |
Db 292 -----ProGlyGlyProAspGlu 297
Qy 307 GGACGAGGAAGTGTGAACCA---TGTCTTAAGGACATTGAGCTGTTTCATGGGAAAGCT 363
| | | | |
Db 298 AspGluGlyProGluGlyProGluGlyProGluGlyGlu---GlyProGluGlyProGlu 316
Qy 364 GGAGAAGGCCCA---GGCAAGACACAGGAGGAAGAAATTTGGGAAAAAACAAGGA 420
| | | | |
Db 317 GlyGluGlyProGluGlyLeuGluGlyProGluGlyGluGlyProGluGlyProGluGly 336
Qy 421 CCA-----GGGAGTCTCAACCCAGGCACA 444
| | | | |
Db 337 ProGluGlyAspSerProAspGlyProAspAlaGlnGluGlyProGluGlyProGlyGly 356
Qy 445 GTACATTGACTCTCCAGAGATCAAGTACAGCTTCAACCTCTGGGAAAGCTGGCCAC 504
| | | | |
Db 357 Pro-----AspGluAspGluGlyProGluGluProGluGlyProGluGlyGlu 372
Qy 505 CTGGTGAAGGAGACAAGTGCCTGAGCTCGTACACATCCTCTTCAAGTCCCTGAACCT 564
| | | | |
Db 373 GlyProGluGly-----ProGluGly 379
Qy 565 CATCTGGCCAGGTGCTCCTGAGGTGGCTAGCAGCCCAAGTGATCTCACCCCTCCTCAC 624
| | | | |
Db 380 ProGluGlyGluGlyProGluGly---ProGluGlyProGluGlyGluGlyProGluGly 398
Qy 625 CCTTAAGCTATCAACCTGCTACAGTCTGTCTAGAGCCACCTGAGAGTAACCTTTGGAT 684
| | | | |
Db 399 -----LeuGluGlyProGluGlyGluGlyProGluGly 409
Qy 685 GGGGTTGGGCCAG-----CCTGGACCACTAGCCGGGCGGACTGGA 725
| | | | |
Db 410 ProGluGlyProGluGlyAspSerProAspGlyProGlyAlaGlnGluGlyProGluGly 429
Qy 726 CAGCGATGAGCCCTGCGCTTACCACCCACATCTCTAGATGACTGGCAACTTCCAGAGC 785
| | | | |
Db 430 -----ProGluGlyProGluGlyGluGlyProGluGlyLeuGluGly 443
Qy 786 CCTCCAGCCAGCACCTTTAGGATACCA-----GGACCTGTTTCCC 827
| | | | |
Db 444 ProGluGlyGluGlyPro-GluGlyProGluGlyProGluGlyGluGlyProGluGlyPr 463
Qy 828 TTCGGCG-----GGGAAGTCATAGTTAGGAGGACCTCACACTTCTCAGGAGAGA 881
| | | | |
Db 463 OgluGlyProGluGlyGluGlyProGluArgProGluGlyProGluGlyGluGlyProGlu 483
Qy 882 CACACAACCATGAGCCTCAGCCTGGGGACCCCACTCCAGGCGCTCCAGCCCAACCTG 941
| | | | |

Db 483 uGlyProGluGlyProGluGlyGluGlyProGlu-----G1 495
QY 942 CCCAGCCAGCCCTGAAATGCAAGTCTGTACGAGTTTGAAGCTAGGAACCCACGGGAAC 1001
Db 495 yProGluGlyProGluGlyGlyGlySer----- 503
QY 1002 TGACTGTGTTCAGGAGAGAGTGGAGTTCTGGACACAGCAAGCGGTGGTGGCTGG 1061
Db 504 -ProAspGlyProGlyAlaGlnGluGlyProGluGlyProGlu-----G1 518
QY 1062 TGAAGTAAGTGGCGGAGGAGCGCTACATTCGAAGCAACATCTGGAGCCCTACAGC 1121
Db 518 yProGluGlyGluGlyProGluGlyLeuGluGlyProGluGlyGlyProGluGlyPr 538
QY 1122 CGGGACCCCTGGGAGCCAGGCGCAGTCACCCCTCTCGGTTCCAAATGCTTCGACTTAGCT 1181
Db 538 oGluGlyProGluGlyGluGlyProGluGlyProGluGlyProGluGly----- 554
QY 1182 CGAGCCCTGAAGAGTTCACAGCTGGCTGCAGGACAGAACTTCTCCACTGCCACGGTGA 1241
Db 555 -GluGlyProGluGlyPro-----GluGlyPr 563
QY 1242 GGACACTGGGTCCTGACGGGAGCCAGCTACTTCGGATAGACTGGGAGCTACAGA 1301
Db 563 oGluArgAspSerProAspGly---ProGlyAlaGlnGluGlyProGluGlyProGluG1 582
QY 1302 TGCTATGTCCACAGGAGGCGCCACCAATCTCTGTC---CCGGTGA---GGCTGTCAAA 1355
Db 582 yProGluGluAspGluGlyProGluGlyProGluGlyProGluGlyGluGlyProGluG1 602
QY 1356 GGATCTGGGTAAGCCCTTAGGCACACAGCTTAGACACCTCCAAGAACCCAGGCCCGCT 1415
Db 602 yProGluGlyGluGlyProGluGlyLeu-----GluGlyProG1 615
QY 1416 GATGCAAGTGCACATCTGATACCATAGAGCCCGCAGAAATCTCTTCTGGATCCCA 1475
Db 615 uGlyAspGluGlyProGluGlyProGluGlyProGluGlyProGluGlyProAspGlyPr 634
QY 1476 GTTTCAGCAAAACCCACACCCAGCCAGCTCACACAGCAAAACCAATGGACAGCCAGAGGC 1535
Db 634 o---GlyAlaGlnGluValProGluGlyProLys-----GlyProGluG1 648
QY 1536 TGAAGCAACAGTGTCCCTTCTGGCTGTGTGA 1569
Db 648 yGluCysGlnSerGlyProSerSerCysGluGly 659
RESULT 10
TI9361
hypotheical protein cl761.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Mar-2000
C:Accession: TI9361
R:White, S.
submitted to the EMBL Data Library, August 1996
A:Reference number: 219114
A:Accession: TI9361
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-925 <WIL>
A:Cross-references: EMBL:278415; PTDN:CAB01670.1; GSPDB:GN00028; CESP:CL761.4
A:Experimental source: clone CL761
C:Genetics:
A:Gene: CESP:CL761.4
A:Map position: X
A:Introns: 40/3; 98/2; 295/1; 443/3; 590/3; 619/3; 691/1; 810/1; 868/3
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology:
Alignment Scores:
Pred. No.: 1.46e-05 Length: 925
Score: 190.50 Matches: 155
Percent Similarity: 33.09% Conservative: 72
Best Local Similarity: 22.59% Mismatches: 202
Query Match: 6.09% Indels: 257

DB: 2 Gaps: 35
us-09-762-021a-1 (1-1710) x TI9361 (1-925)
QY 22 CCTGCAGAGGCTCTGGAGGAGAGCTGGAGCAAGAGACTCGACTTGGAGGCTTCAGCC 81
Db 46 ProAlaAsnGlyTyrGlyGlyMetProGlyPro-ProAsnIleGlyGlyProGlnPr 65
QY 82 A-----GGCAGGACAGATGGAGGGCGCTGCTATGAAAG 117
Db 65 oProGlyGlySerGlnGlnLysGlyGlnGlnGlnPheProGlySerGlyAlaAsnMe 85
QY 118 CGCGTCCCT-----ATGAGCAGGACACGCTATCTGGAGCCGGGAT 159
Db 85 tLysLeuProGlyTyrAspGlyAsnSerMetGlnAsnAla---TyrMet---ProTyrPr 103
QY 160 CCCTCCCAAC-----AGCCCAACAGAGGACCCCTAGACACAGCT 201
Db 103 oPro-GlnAsnGlnArgSerGlyGlyGlnAlaProGlnAsnGlyProProAsnTyrA 123
QY 202 CCCACATCCCAAGGCCCTGCCACGACACAGCTGCCGAGAACCA-----G 252
Db 123 snSerHisGlnGlnMetPro-----ProAsnAsnGlnTyrGlyGlyV 137
QY 253 TGCCT-----TTACTCTGCCTCTCCCAAGGGGCTCT---CTTCCCGCAGGA 297
Db 137 alProAspProTyrArgMetTyrProGlyMetGlnGlyProGlyGlnValProAsnS 157
QY 298 CCCAGAGGAGGAGGAGGAGTGTGAACCATGTCTTAAGGCACATGAGCTTTCATGG 357
Db 157 erGlnAsnSerSer----- 161
QY 358 AAAGCTGGAGAGGCC-----AGCAAGACACAGCAGGAAGAAGAA 399
Db 162 --SerGlnGlnArgProProSerGlnAsnAlaAsnGlnGlnArgProAlaSerGln---- 179
QY 400 ATTTGGGAAAAAACAAGGAGGAGTCTCACCCAGGAGGCTGGCCACCTGGCTGAAGAGAC 459
Db 180 --AlaGlyGlnGlnTyrProThrGlnGlnAlaLeuProAlaHis----- 193
QY 460 CCAGAGATCAAGTACAGCTTCAACCTCTGGGAGGCTGGCCACCTGGCTGAAGAGAC 519
Db 194 -----LeuHisGlyThrProThrTyrProGlyMetProPro-----GlnA 207
QY 520 AAGTCCCTCGAGCTCGTACATCTCTTCAAGTCCCTGAACCTCATCTGCGCCAGGTG 579
Db 207 snAlaProLeuGlnHisTyrGln----- 214
QY 580 CCCTGAGGCTGGCTTAGCAGCCCAAGTATCTACCCCTCTCACCCTTAAG-----C 633
Db 215 -----ProGlnTyrAlaAsnGlyThrProProValArgGlyAlaG 229
QY 634 TATCAACCTGTACAGTCTCTTAAGCCACCTCAGAGTACCTTTGGATGGGTTGGG 693
Db 229 lySerSerAlaPheProProLeu----- 236
QY 694 CCCAGCTGGACCACTAGCCGGCGGCGACTGGACAGCGCATGAGCCCTGCCCTTACCAAC- 752
Db 237 --GlnProSerLysGlnSerLysProAspGluHisArg-----ProAsnAsnL 252
QY 753 --CCACATTCACAGTACTGGCAACTTCCAGAGCCCTCCAGCCCAAGCACCCTTAGGATA 810
Db 252 euSerAsnSerGlnTyrProGlyAsnPheGlyAlaProGlySerSerSerGlyPheAspS 272
QY 811 CCAGACACCTGTTTCCCTTCGGCGGGGAGTATAGTTAGGAGCACCCTCACACTTCC 870
Db 272 erPheSerAsnGlySerSerGlyTyrGln-----GlyTyrGlyLeuPro----- 286
QY 871 TCAGGAGAGACACACCAACCATGACCCCTCAGCCTGGGACCCCAACTCCAGCCCTCCAG 930
Db 287 -----GlySerGlyThrProGlySerGlnG 295
QY 931 CCCCAAACCTGCCCCAGCCAGCCCTGAAATGCAAGTCTTGTACGAGTTTGAAGCTAGGAA 990


```
Db 295 lyHisSerThrSerGlyGlnAsnThrSerAlaGlnAsnAlaThrSerGlyLeuThrGly- 314
    :::      |||      :::::  |||||
Qy 991 CCACGGGAACGTGCTGGTCCAGGAGAGAGCTGCAGGTTCT- 1035
    ||||| |||      |||||:::
Db 315 ProThrGlnAsnProPheGlySerMetAsnGlyGlnGlyAsnHisGlyGlnPheGly 334
    |||||::: |||      |||||
Qy 1036 GGACCACAGCAAGCGGTGGTGGTGAAGAAATGAGCGGAGCGGCTACATGCC 1095
    |||      ::: ||| |||||
Db 335 GlyAsnAspLeuSerGlyValProGlyProSer---GlyGlyTyrGlyGlnMetAsnSer 353
Qy 1096 -----AAGCAACATCTGGAGCCCT-ACAGCCGGGAC 1128
    ||||| ||||| ||||| |||||
Db 354 SerAsnThrProAsnGlnSerGluArgSerThrProGlyGlnProSerThr-ProGlyTh 373
Qy 1129 CCTCGGACCCAGCGGAGCTCACCTCTCGGTTCCATGCTTCGACTAGCTCGAGGCC 1188
    |||||::: ||| ||||| |||||
Db 373 rProGlySerHisGlyThrValGlySerAsnAlaProLeu-----SerHisG1 389
Qy 1189 TGAAGAGGTCACAGACTGGCTGCAGGCAGAGAACTTCCACTGCCAGCGTGAGGACACT 1248
    ||| ||||| ::|||::: ||||| |||
Db 389 n-LysProSerGln-----GlnGlnGlnSer-MethHisAsnHis----- 401
Qy 1249 TGGTCCCTGACGGGAGCCAGCTACTTCGCATATAGACCTGGGAGCTACAGATGCTATG 1308
    ||| ||||| |||      :::
Db 402 -----LeuProHisHisGlnTyrAsnGlnAsnLeu 413
Qy 1309 TCCACAGGAGGCCCAAGAACTCTCCCGGCTGGAGCTGTGCAGAAAGATGCTGGGAT 1368
    ||| |||      ::: ||||| |||||
Db 414 Ser-----ProAsnHisGlyAlaSerSerLeuGlySerGlnGlyHisGlySer 430
Qy 1369 AAGCCCTTAGGCACAGCTTAGACACCTCCAAGAACCCGCGCTGATGCAAGATGCC 1428
    |||||::: ||||| ||||| |||
Db 431 SerProMetGlySerSerLeuMetProLeuAsnGlyGlnTyrProSerMetThrGlnAsn 450
Qy 1428 ----- 1428
Db 451 MetGlnSerProAlaSerThrSerMetGluProThrPheLysGluProAlaValProIle 470
Qy 1429 -----AGATCTGATACCCATTAGAGCCCGCCGAG 1455
    ||||| ||||| ||||| |||||
Db 471 ArgHisSerProSerGlnMetProThrHisLeuGlnSerProValHisProSerProAsn 490
Qy 1456 -----AATTCCTCTTCTGGATCCCGAGTTTGCAGCAAAACCCACACCC 1497
    ||||| ||| |||||
Db 491 GlyAlaProAlaTyrAsnAlaProSerSerLys-----ThrPro 505
Qy 1498 CAGCTCACACGCAA----- 1512
    ||||| |||||
Db 506 AspProThrGlnGlnArgProHisSerProThrPheAlaValProThrLeuProAla 525
Qy 1513 -----AAACAATGGACAGGC 1527
    ||||| |||
Db 526 AlaAlaThrLeuAlaGlnAlaPheSerAlaAsnGlnIleSerThrLysProLysThrSer 545
Qy 1528 CCAGAGGCTGAA-----GCAAAAGTGTCCCTTCTGGCTGTGTGGAGCTCCCGACGTA 1581
    ||||| ::: ||||| |||||
Db 546 ProGlnLysLysHisHisGluAspGlyValProGluProProThrAlaAspThrProPhe 565
Qy 1582 ACCACCTATTATTATTACTCTTTCCCAAACTGGAGCATTTATG----- 1626
    ||||| ||| |||||
Db 566 ThrThrValThrHisTyrGluLeuProAlaAlaMetThrPheLeuArgAspThrLeuHis 585
Qy 1627 -----CCTAGGCTTGTCAAGAATCTGTTGAGTCCCTCTCTCTCTCTCTCTCT 1662
    |||||::: ||||| |||||
Db 586 ValGlyProAsnAspLysValHisProGlnValGluLysHisTyrPheSerArgLysArg 605
Qy 1663 TCTCAA 1668
    |||
Db 606 GlnGln 607
RESULT 11
A54849
```

```
collagen alpha 1(VII) chain precursor - human
N:Alternate names: procollagen alpha 1(VII) chain
C:Species: Homo sapiens (man)
C:Date: 04-Nov-1994 #sequence_revision 04-Nov-1994 #text_change 20-Sep-1999
C:Accession: A54849; PH0844; S16316; I56328; A30296; I84686
R:Christiano, A.M.; Greenspan, D.S.; Lee, S.; Uitto, J.
J. Biol. Chem. 269, 20256-20262, 1994
A:Title: Cloning of human type VII collagen. Complete primary sequence of the alpha1(
A:Reference number: A54849; MUID:94327588; PMID:8051117
A:Accession: A54849
A:Status: not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2944 <CHR>
A:Cross-references: GB:L02870; NID:g987124; PIDN:AAA75438.1; PID:g987125
R:Tanaka, T.; Takahashi, K.; Furukawa, F.; Imamura, S.
Biochem. Biophys. Res. Commun. 183, 958-963, 1992
A:Title: Molecular cloning and characterization of type VII collagen cDNA.
A:Reference number: PH0844; MUID:92231902; PMID:1567409
A:Accession: PH0844
A:Molecule type: mRNA
A:Residues: 'EPR',340-475,'RALSTASHSTLCWRATRHPCNRGSHWTRAAACEPCNRPAASHRAARAG',524-528,'
A:Cross-references: DDBJ:D11152; DDBJ:D13694; NID:9453698; PIDN:BAA02853.1; PID:94536
A:Experimental source: keratinocyte
R:Parente, M.G.; Chung, L.C.; Ryyanen, J.; Woodley, D.T.; Wynn, K.C.; Bauer, E.A.;
Proc. Natl. Acad. Sci. U.S.A. 88, 6931-6935, 1991
A:Title: Human type VII collagen: cDNA cloning and chromosomal mapping of the gene.
A:Reference number: S16316; MUID:91334380; PMID:1871109
A:Accession: S16316
A:Molecule type: mRNA
A:Residues: 815-892,'E',894-1439 <PAR>
A:Cross-references: GB:M65158; GB:S49017; NID:g180914; PIDN:AAA96439.1; PID:g180915
A:Experimental source: keratinocyte
R:Gammon, W.R.; Abernethy, M.L.; Padilla, K.M.; Prisayanh, P.S.; Cook, M.E.; Wright,
J. Invest. Dermatol. 99, 691-696, 1992
A:Title: Noncollagenous (NC1) domain of collagen VII resembles multidomain adhesion p
A:Reference number: I56328; MUID:93107742; PMID:1469284
A:Accession: I56328
A:Status: translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 'EPR',372-517,'DV',520-540,'W',542-1255 <RES>
A:Cross-references: GB:S51236; NID:9262308; PIDN:AAB24637.1; PID:g262309
R:Seltzer, J.L.; Eisen, A.Z.; Bauer, E.A.; Morris, N.P.; Glanville, R.W.; Burgeson, F.
J. Biol. Chem. 264, 3822-3826, 1989
A:Title: Cleavage of type VII collagen by interstitial collagenase and type IV colla
A:Reference number: A30296; MUID:89139437; PMID:2537292
A:Accession: A30296
A:Molecule type: protein
A:Residues: 'A',1240-1246,'G',1248-1250,'XE',1253-1255,'Q',1257,'E',2032,'C',2034-204
A:Note: two reported peptides cannot be reliably located
R:Greenspan, D.S.
Hum. Mol. Genet. 2, 273-278, 1993
A:Title: The carboxyl-terminal half of type VII collagen, including the non-collagenou
A:Reference number: 148103; MUID:93271985; PMID:8499916
A:Accession: I84686
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: mRNA
A:Residues: 2395-2871,'S',2873-2944 <PE2>
A:Cross-references: GB:L06862; NID:g388713; PIDN:AAA89196.1; PID:g388714
R:Christiano, A.M.; Ryyanen, M.; Uitto, J.
Proc. Natl. Acad. Sci. U.S.A. 91, 3549-3553, 1994
A:Title: Dominant dystrophic epidermolysis bullosa: identification of a Gly --> Ser is
A:Reference number: A55255; MUID:9422477; PMID:8170945
A:Contents: annotation
C:Comment: Prolines and lysines at the third position of the tripeptide repeating uni
ed and subsequently O-glycosylated.
C:Genetics:
A:Gene: GDB:COL7A1; EBR1; EBD1; EB
A:Cross-references: GDB:128750; OMIM:120120
A:Map position: 3p21.3-3p21.3
A:Note: defects in this gene can result in dominant and recessive dystrophic epidermo
A:Note: there are 118 introns
C:Complex: type VII collagen is probably a homotrimer
```

C:Function:

A:Description: structural component of extracellular polymer associated with anchoring of
 C:Superfamily: unassigned collagens; animal kunitz-type proteinase inhibitor homology;
 C:Keywords: coiled coil; extracellular matrix; glycoprotein; hydroxylysine; hydroxyproline
 F:1-16/Domain: signal sequence #status predicted <SIG>
 F:17-2944/Product: collagen alpha 1(VII) chain #status predicted <MAT>
 F:17-1253/Domain: amino-terminal nonhelical #status predicted <NC1>
 F:36-201/Domain: von Willebrand factor type A repeat homology <VWA1>
 F:321-318/Domain: fibronectin type III repeat homology <FN1>
 F:327-413/Domain: fibronectin type III repeat homology <FN2>
 F:414-502/Domain: fibronectin type III repeat homology <FN3>
 F:508-593/Domain: fibronectin type III repeat homology <FN4>
 F:598-683/Domain: fibronectin type III repeat homology <FN5>
 F:686-771/Domain: fibronectin type III repeat homology <FN6>
 F:776-862/Domain: fibronectin type III repeat homology <FN7>
 F:864-952/Domain: fibronectin type III repeat homology <FN8>
 F:954-1045/Domain: fibronectin type III repeat homology <FN9>
 F:1052-1219/Domain: von Willebrand factor type A repeat homology <VWA2>
 F:1170-1172/Region: cell attachment (R-G-D) motif
 F:1189-1253/Region: cysteine/proline-rich
 F:1254-2783/Region: interrupted helical
 F:1334-1336/Region: cell attachment (R-G-D) motif
 F:2008-2010/Region: cell attachment (R-G-D) motif
 F:2553-2555/Region: cell attachment (R-G-D) motif
 F:2784-2944/Domain: carboxyl-terminal nonhelical #status predicted <NC2>
 F:2876-2929/Domain: animal kunitz-type proteinase inhibitor homology <BPI>
 F:337-786,1109/Binding site: carboxylate (Asn) (covalent) #status predicted
 F:2167,2176,2185,2188,2664,2667,2673/Modified site: 4-hydroxyproline (Pro) #status exper
 F:2625,2631/Modified site: 5-hydroxylysine (Lys) #status experimental
 F:2625,2631/Binding site: carboxylate (Lys) (covalent) #status experimental
 F:2634,2802,2804/Disulfide bonds: interchain #status predicted

Alignment Scores:

Pred. No.:	1-76e-05	Length:	2944
Score:	190.00	Matches:	144
Percent Similarity:	32.3%	Conservative:	27
Best Local Similarity:	25.71%	Mismatches:	322
Query Match:	6.08%	Indels:	158
DB:	2	Gaps:	31

US-09-762-021A-1 (1-1710) x A54849 (1-2944)

QY	1	GGCAGGCGACTGAGACAGCGCTGTGAGAGGCTCTGGAGGAGAGCTGGAGCAAGACC	60
Db	1278	GlyArgThrGlyAlaProGlyPro--GlnGlyProProGlySerAlaThrAlaLysGly	1296
QY	61	TCGACTTGGAGCCCTCAGCC-----AGCCAGGACACATGGAGGGGCGCTGTATGA	114
Db	1297	GluArgGlyPheProGlyAlaAspGlyArgProGlySerProGlyArgAla-----Gly	1314
QY	115	AAGGCGCGCTTATGGAGCGCAGCGCTATCTGGAGCGGGGATCCCTCCAGAACAGCC	174
Db	1315	AsnProGlyThrProGlyAlaProGlyLeuLysGlySer-----ProGlyLeuProGly	1332
QY	175	GCACAGAGAGCCCTAGACAGAGCTCCACCATCCCAAGGCC-----CCTGCCAGGCCA	231
Db	1333	ProArgGlyAspProGlyGluArgGlyProArgGlyProLysGlyGluProGlyAlaPro	1352
QY	232	CACCACTGCCCGAGAACCAAGTGCCCTTACTCTGCCTCTCCCAAGCGGCTCTCTCC	291
Db	1353	GlyGlnValIleGlyGlyGlyProGlyLeuProGlyArgLysGlyAspProGlyPro	1372
QY	292	CGAGGACCCAGAGGAGGAGGAGAGT-----GCTGAACCATGCTCCCAAGGACATTGAGCT	348
Db	1373	SerGlyProProGlyProArgGlyProLeuGlyAspProGlyProArgGlyProProGly	1392
QY	349	GTTTCATGGGAAGCTGGAGAGGCCAGCCAGCAAGACCCAGCAGAGAGAAATTTGGGAA	408
Db	1393	LeuProGlyThrAlaMetLysGlyAspLysGlyAspArgGlyGlu-----Arg	1408
QY	409	AAAAACAGAGGAGGAGGCTCAC-----	435
Db	1409	GlyProProGlyProGlyGlyGlyGlyIleAlaProGlyGluProGlyLeuProGlyLeu	1428

QY	436	CCAGGCACAGTACATTGACTCTCCAGAAGATCAAGTACAGCTTCAACCTCTCTGGGAAG	495
Db	1429	ProGlySer-----ProGlyProGlnGlyProValGlyProProGlyLys	1443
QY	496	GCTGCCACCTGGCTGAAGGAGACAGTGCCCTGAGCTGCTGACATCTCTTCAAGTC	555
Db	1444	LysGlyGluLysGlyAspSerGluAsp-----GlyAla	1454
QY	556	CCTGAACATTCCTCCCTGGCAGCTGCCCTGAGCTGGCCTAGCAGCCCAAGTATCTCACC	615
Db	1455	ProGlyLeu---ProGlyGln-----ProGlySerProGlyGluGlnGly	1468
QY	616	CCTCTCATCCCTAAAGCTATCAACCTGCTACAGTCTCTTAAGCCCACTGAGAGTAA	675
Db	1469	ProArgGlyPro-----ProGly-----	1474
QY	676	CCTTTGGATGGGTGGGCCAG-----CCTGGACCACTA-----GCC	713
Db	1475	-----AlaIleGlyProLysGlyAspArgGlyPheProGlyProLeuGlyGluAla	1491
QY	714	GGGCGACTGGACAGCGATGAGCCCTGCTACCAACCCACATCTCAGATGACTGCG	773
Db	1492	GlyGluLysGlyGluArgGlyProProGlyProAlaGlySerArgGly---LeuProGly	1510
QY	774	AACTTCCAGAGCCCTCCAGCCCAAGCAGCTTAGGATACAGGACCTGTTCCCTTCCGC	833
Db	1511	ValAlaGlyArgProGlyAlaLysGlyPro-Glu-----GlyProProGlyProThrG	1528
QY	834	GGGGAAGTCATAGTTAGGAGCACCTCACATTTCTCAGGAGAGACACACACCA--	891
Db	1528	yArgGlnGly-----GluLysGlyGluProGlyArgProG	1540
QY	892	TGACCTCAGCTGGGAGCCCAACTC-----CAGGCCCTCCAGCCCAACACCG	941
Db	1540	yaspProAlaValValGlyProAlaValAlaGlyProLysGlyGlyValGlyAspValG	1560
QY	942	CCCAGCAGCCCTGAAAATGCAAGTCTTGTACGAGTTTGAAGCTAGGAACCCACGGAA	1001
Db	1560	yProAlaGlyProArgGlyAlaThrGlyValGln-----	1571
QY	1002	TGACTGTGTCAGGAGAGAGCTGGAGGTTCTGGACCACAGCAAGCGGTGGTCTCG	1061
Db	1572	-----GlyGluArgGlyProProGlyLeuValLeuProG	1583
QY	1062	TCAAGAATGAGCGGAGCGGCTACATTTCCAAGCAACATCTCGAGGCCCTACAGC	1121
Db	1583	yasp-----ProGly-----Proly	1588
QY	1122	CGGGACCCCTGGAGCCAGGCCAGTCACTCTCGGTTCCAATGCTTCGACTTAGCT	1181
Db	1588	sglyAspProGlyAspArgGlyProIleGlyLeuThrGlyArgAlaGlyProProGlyAs	1608
QY	1182	CGAGCCCTGAAGAGTTCACAGACTGGCTGCAGGCAGAGAA-----CTTCTCCA	1229
Db	1608	pserGlyProProGlyGlyLysGlyAspProGlyArgProGlyProProGlyProValG	1628
QY	1230	CTGCACCGGTGAGCACCTTGGTC-----CCTGACGGGAGCC	1268
Db	1628	yProArgGlyArgAspGlyGluValGlyGlyLysGlyAspGlyGlyProProGlyAspP	1648
QY	1269	AGCTACTTGCATAGACCTGGGAGCTGTGAGAGGATGCTCCAGAGGAGGCCACGAA	1328
Db	1648	oglyLeuProGlyLysAla---GlyGluArgGlyLeuArgGlyAlaProGlyValThrG	1667
QY	1329	TCCTGTCGGCTGGAGGCTGTGAGAGGATGCTGGGATAGCCCTTAGGCACCACTT	1388
Db	1667	yProValGlyGlyLysGlyAspGlnGlyAspProGly-----	1679
QY	1389	AGACACCTCCAAGAACCGGCCCGCTGTGATGCAAGATGCGAGTCTGTATACCATTAG	1448
Db	1680	-----GluaspGlyArgAsnGly-----Se	1686

QY 1449 CCCCAGAAATTC-----TCTCTGGATCCAGTTTGCAGCAAAACCCACACACC 1496
| | | | : : : | | |
Db 1686 rProGlySerSerGlyProLysGlyAspArgGlyProGlyProGlyProGlyProGly 1706
QY 1497 CCAGCTCACACAGCAAAACAATGACACAGCCAGGAGGCTGAACCAACAGTGTCCCT 1554
: : : | | | | | | | | | |
Db 1706 yArgLeuValasp-----ThrGlyProGlyAlaArgGlyLysGlyGluPro 1721
RESULT 12
S49915
extensin-like protein - maize
C:Species: Zea mays (maize)
C:Date: 05-Mar-1995 #sequence_revision 12-May-1995 #text_change 29-Oct-1999
C:Accession: S49915
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.
submitted to the EMBL Data Library, June 1994
A:Description: Pex genes: pollen-specific genes with extensin-like domains.
A:Reference number: S49915
A:Accession: S49915
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1188 <RUB>
A:Cross-references: EMBL:Z34465; NID:g600117; PIDN:CAA84230.1; PID:g600118
Alignment Scores:
Pred. NO.: 1.91e-05 Length: 1188
Score: 189.00 Matches: 140
Percent Similarity: 31.56% Conservative: 50
Best Local Similarity: 23.26% Mismatches: 235
Query Match: 6.04% Indels: 177
Gaps: 23
US-09-762-021A-1 (1-1710) x S49915 (1-1188)
QY 79 GCCAGGCCAGGACAG---ATGGAGGGCCCTGCTATGGAAGGCGCTCCCTATGGAGCA 135
| | | | | : : : | | | | |
Db 396 AlaArgProValaspCysSerLysHisValCysAlaGlyTyrProThrPro---GlyGly 414
QY 136 GGCACGCTATCTGGA-----GCCGGGATCCCTCCAGACAGCCGCCACAGAGGACCT 189
| | | | | | | | | | : : : | | | | |
Db 415 GlyProProSerSerProValProGlyLysProAlaAlaSerAlaProMetProSerPro 434
QY 190 AGAGCACAGCTCCACACATCCCAAGGCCCTGCCAGCCGACACACAGTGCCTCCGAGAAC 249
| | | | | | | | | | | | | | | |
Db 435 HisThrProProAspVal-SerProGluProLeuPro-----GluPr 448
QY 250 AAGTGCCTTTACTTGCTCTCCTCAAGGCGGTCTCTTCCGCCAGGACCCAGAGAGGA 309
| | | | | | | | | | : : : | | | | |
Db 448 oSerProValProAlaProAlaProMetProMetProThrProHisSerProAlaAs 468
QY 310 CGAGGAAGTGTGAACCATGCTCCTAAGGGACATTGAGCTGTTCATGGAAAGCTGGAGAA 369
| : : | | | | | | | | | |
Db 468 pAspTyrVal--ProProThrPro----- 475
QY 370 GCCCCAGGCAAGACACAGGAGAGAGAAATTTGGGAAAAAACAAGGACCGAGGAGG 429
| : : | | | | | | | | | | : : : | | | | |
Db 476 -----ProValProGlyLys 480
QY 430 TCTCACCCAGGCACAGTACATTGACTGCTTCCAGAGATCAAGTACAGCTTCAACCTCCT 489
| : : | | | | | | | | | | : : : | | | | |
Db 481 SerProAlaThr-----SerProSerProGlnValGlnProPro 494
QY 490 GGGAAAGGTGGCCACCTGGCTGGAAGGAGACAAGTGGCCCTGAGCTCGTACACATCCTCT 549
| : : | | | | | | | | | | : : : | | | | |
Db 495 Ala-----AlaSerThrProProPro 501
QY 550 CAAGTCCCTGAACCTCATCTCGGCCAGGTGGCCCTGAGGCTGGCTAGCAGCCCAAGTGCAT 609
: : : | | | | | | | | | | : : : | | | | |
Db 502 SerLeuValLysLeuSerProProGlnAlaProValGlySerProProProValLys 521
QY 610 CTCACCCCTCCT-----CACCCCTAAAGCTATCAACCTGTACAGTCTCTG 654
| | | | | | | | | | : : : | | | | |
Db 522 ThrThrSerProAlaProIleGlySerProSerProProProValSerValVal 541

QY 655 TCTAAGCCCACTGAGAGTAACCTTTGGATGGGTTGGGCCCACTGGACCACTAGCCG 714
| | | | | : : : | | | | | : : : | | | | |
Db 542 SerProProPro-----ProValLysSerProProPro--ProAlaProValGlyS 558
QY 715 GCGCGACTGGACAGCGGATGAGCCCTGCGCTTACCACCCACATCTCTAGATGACTGC 774
| | | | | | | | | | | | | | | |
Db 558 erProProProGluLysSerProProProAlaProValAlaSerProProProp 578
QY 775 ACTTCCAGAGCCCTCCAGCCCAAGCACCTT----- 804
: : : | | | | | | | | | | : : : | | | | |
Db 578 roValLysSerProPro--ProProThrLeuValAlaSerProProProProValLysSer 597
QY 805 -----AGGATACCAGGACCTGT 822
Db 598 ProProProProAlaProValAlaSerProProProValLysSerProProPro 617
QY 823 TTCCTTCGCGGGGGAAGTCATAGTTAGGAGGACCTCACACTTCTCTCAGGAGAGAC 882
| | | | | | | | | | : : : | | | | |
Db 618 ThrProValAlaSerProPro-----ProProAlaProValAlaSerProPro 634
QY 883 ACACAACCATGACCTCAGCTCAGCTGGGACCCCACTCCAGGCCC-----TCCAG 930
| | | | | | | | | | : : : | | | | |
Db 635 ProMet-LysSerProProProProThrProValSerSerProProProGluLysSe 654
QY 931 CCCCAAACTGCCAGCCAGCCCTGAAAAATGCAAGTCTTGTACGAGTTTGAAGTAGGAA 990
| | | | | | | | | | : : : | | | | |
Db 654 rProProProProProAlaLysSerThrProProGluGluTyr----- 670
QY 991 CCCACGGGAAGTACTGTGTGTCAGGAGAGAGCTGGAGTTCTGGACCAACAGACG 1050
| | | | | | | | | | : : : | | | | |
Db 671 -ProThrProThr----- 675
QY 1051 GTGTGTGCTGCTGAAGATGAGCGGGACGGCGCTTACATTCACAGCAACATCCTG-- 1108
| | | | | : : : | | | | | : : : | | | | |
Db 676 -----SerValLysSerProProProGluLysSerLeuProProThrLeu 693
QY 1109 -----GAGCCCTACAGCCGGGACCCCTGGGACCCAGGCGCA 1146
: : : | | | | | | | | | | : : : | | | | |
Db 693 eProSerProProProGlnGluLysProThrProProSerThrProSerLysProProSe 713
QY 1147 GTCACCTCTCGG-----GTTCCAATGCTTCGAGTTAGTCGAGGCTGAAGAGGTCA 1200
| | | | | : : | | | | | : : | | | | | : : | | | | |
Db 713 rSerProGluLysProSerProLysGluProValSerSerProProGlnThrProLy 733
QY 1201 AGACTGGCTCAGCAGCAGAACTTCTCCACT-----GCCACGGTCAG 1242
| | | | | : : : | | | | | : : : | | | | |
Db 733 sSerSerProProProAlaProValSerSerProProProThrProValSerSerProPr 753
QY 1243 GACACTTGGTCCCTGACGGGGAGCCAGCTACTTCGCATAAGACCTGGGAGCTACAGAT 1302
| | | | | : : : | | | | | : : : | | | | |
Db 753 oAlaLeuAlaProValSerSerProProSerValLysSerSerProProAlaProLe 773
QY 1303 GCTATGTCCAGAGGAGGCGCCCAAGATCTGTCTCCGGCTGGAGGCTGTCAAGAGGTCT 1362
| | | | | | | | | | : : | | | | | : : | | | | |
Db 773 userSerProProProAlaProGlnValLysSerSerProProProValGln-----Va 791
QY 1363 GGGGATAAGCCCTTAGGCACAGCTTAGACACCTCAAGAACACCGCCCGCTCATCAA 1422
| : : | | | | | | | | | | : : : | | | | |
Db 791 lSerSerProProProAlaProLysSerSerProPro----- 803
QY 1423 GATGGCAGATCTGTATCCATTTAGAGCC-----CCGAG 1455
| | | | | : : : | | | | | : : : | | | | |
Db 804 -----LeuAlaProValSerSerProProGlnValGluLysThrSerProProPr 820
QY 1456 AATTCTCTTCTGGATCCCATTTTGCAGCAAAACCCACAGCTTACACAGCAAAAA 1515
| | | | | | | | | | : : | | | | | : : | | | | |
Db 820 oAlaProLeuSerSerProProLeu-AlaProLysSerSerSerProHisValVal--V 839
QY 1516 CAATGGACAGCCAGAGGCTGAAGCAACAGTCTCTCTCTGCTGTGTGGAGCTCC 1575
: : : | | | | | : : : | | | | | : : : | | | | |
Db 839 alSer-SerProProProValValLysSerSerProProProAlaProValSerSerPro 858


```
QY 1348 TCTCAGAAGATGCTGGGGATGAAGCCCTTAGGCACCAAGCTTAGACACCTCCAAGAACCCAG 1407
Db 505 ---GlnGlnValMetArgProThrProSerProThrGlySerSerGlySerArgSerMet 523
QY 1408 GCCCGCTGTATGCAAGATGGCAGATCTGATACCCATAGAGCCCGCAGAAATTCCTCTCT 1467
Db 524 SerProAlaValAlaGlnAsnHisProIleSerArgProAlaSerAsnGlnSerSerSer 543
QY 1468 GGATCCCAGTTT-----GCAGCAAAACCCACACCC----- 1497
Db 544 GlyGlyProMetGlnGlnProProValGlyAlaGlyGlyProProMetProProHis 563
QY 1498 -----CAGCTCACACAGCAAAACCAATGGACAGGC 1527
Db 564 ProGlyMetProGlyGlyProGlnGlnGlnSerGlnGlnGlnAlaSerAsn 583
QY 1528 CCAGAGGCTGAAGCAACAGTCTCCTCTGCTGCTGTGGAGCCTCCCGAGTAACCAACC 1587
Db 584 SerAlaSerSerAlaSerAsnSerProGlnGlnThrProProAlaProPro----- 601
QY 1588 TATTATTATTACCTTTCCCAACCTCGAGCATTTATGCGCTTAGCCTTGTCAAGAACTCG 1647
Db 602 -----ProasnGlnGly-----MetAsnAsnMet 609
QY 1648 TTCAGTCCCTCTCTCT 1662
Db 610 AlaThrProProPro 614
RESULT 14
T29074
Hypothetical protein SC1C2.25c - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 11-May-2000
C:Accession: T29074
Mol. Microbiol., M.: Kieser, H.M.; Denapalite, D.; Eichner, J.; Cullum, J.; Kinashi, H.; Hopwood, J.A.
A:Title: A set of ordered cosmid clones and a detailed genetic and physical map for the 8 Mb S. coelicolor chromosome.
A:Reference number: 220556; PMID:97000351; PMID:8843436
A:Accession: T29074
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1329 <RED>
A:Cross-references: EMBL:AL031124; NID:e1312893; PID:e1312918; PIDN:CAAL19992.1
C:Genetics:
A>Note: SC1C2.25c
Alignment Scores:
Pred. No.: 3.39e-05 Length: 1329
Score: 185.50 Matches: 139
Percent Similarity: 31.15% Conservative: 51
Best Local Similarity: 22.79% Mismatches: 220
Query Match: 5.93% Indels: 201
Db: 2 Gaps: 27
US-09-762-021A-1 (1-1710) x T29074 (1-1329)
QY 19 CAGCCTGCAGAAAGCTCTGGAGGAAGAGCTGGAGCAAGAACCTCGACTTGGAGGCTTCA 78
Db 725 GlnProGlyAlaGlyAsnGlyGlyProAlaAlaGlnGlnAlaAlaGlyValAlaAla 744
QY 79 GCCAGCCAGGACAGATGGAGGGCGCTGCTATGGAAAGCGCGCTCCCTATGGACGAGC 138
Db 745 AlaArgArgGlyGlnGlyGlyGlyAlaLeuAlaGlyGlyGlyAlaPheGlyGlyGly 764
QY 139 AG-----CTATCTGGAGCCGGGATGCC-----TCAGAACACCCCGCAGAGAC 186
Db 765 SerGlyGlyGlyAlaLeuGlyAlaGlyAlaProGlyGlyArgLeuGlyAlaGlyGlnGly 784
QY 187 CCTAGACGACACCTTCCCAACCATCCCAAGGCCCTGCCACGCCACACCATGTCGCCGAGA 246
Db 785 ProArgAlaAlaLeuPro----- 790
```

```
QY 247 ACCAAGTGCCTTTACTGCTCTCTCCAAGGGCGTCTCTTCCCGGAGGACCCAGAGAG 306
Db 791 -----GlyArgAspAlaGlyGlyArgProGlyAlaProGlyGly 803
QY 307 GGACGAGGAAGTGTGAACCATGTCTTAAGGACACATTGAGCTGTTTCATGGGAAACCTGGA 366
Db 804 AlaArgGlyProGlnSerProAlaValProGlyGlnGlnAsnArgProAlaProAlaGly 823
QY 367 GAAGCCCGCAGGCAAGACGACGAGCAAGAAATTTGGGAAAAAACAAGACCCAGG 426
Db 824 AlaGlyAlaGlyAlaGlyGlnGlnAlaProGlyAlaProGlyGlnGlnAlaAlaGly 843
QY 427 AGGTCTCACCCAGCAGCAGTACATTGCTTCCAGAGATCAAGTACAGCTTCAACCT 486
Db 844 MetAsnAlaPro-----GlnGly 849
QY 487 CTGGGAAGGCTGGCCACTGGCTGAAGGACACAAGTGCCTCGAGCTGTACACATCTCT 546
Db 850 AsnAlaProAlaAsnAlaSerGlnGlyGlnAspMetPheGlyAsnArgAlaPro 869
QY 547 CTTCAAGTCCCTGAACTTCATCTCGCCAGGTGCTGAGGCTGCCTAGCAGCCCAAGT 606
Db 870 GlyAlaProProGlnArgLysArgGlyGly-----Asn-GlyAspAlaGlnGln 886
QY 607 GATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGCTACAGTCTCTGTCTAAGCCACC 666
Db 886 yArgArgProGlnLeuProProArgGly----- 895
QY 667 TGAGAGTAACTTTGGATGGGTGGGCCACCCCTGGACCCTAGCCGGGCGACTGAC 726
Db 896 -----GlyPro-----ArgAlaGlnLeuPr 902
QY 727 AGGCGATGAGCCCTGCCCTACCAACCCACATCTCAGATGAC-----TGCAACT 777
Db 902 oGlyGlyAsnProGlnProArgValProSerTrpSerAspGluAsnAlaGlnProProVa 922
QY 778 TCCAGAGCCCTCCAGCAAGCACCTTAGTATACAGCACCCCTGTTTCCCTTCGGCGGG 837
Db 922 lProArgAlaSerLeuAspAlaProArgGlyHisAspGluProAspSerSerArgThrAs 942
QY 838 A---AGTCATAGGTTA-----GGGAGCACCTCACACTTTCCTCAGGA 876
Db 942 pArgThrProArgLeuAspGluArgGlnGlyProGlySerThrAlaGluMetProAlaVa 962
QY 877 GAAGACACACACCATGACCTCAGCTGCGGACCCCAACTCC----- 919
Db 962 l-----ProArgPheGlyGluProGlnSerProAlaAlaThrAlaGln 976
QY 920 -----AGGCCC-----TCCAGCCCCAACCT----- 940
Db 976 uPheAlaArgProAspPheAspAlaProAlaProArgArgAspGluSerGlnAspThrGln 996
QY 941 -----GCCAGCCAGCCCTGAAATGCAAGCTTGTACGAGTTTGAAGCTAGGAACCC 993
Db 996 yGlnThrAlaGlnProGlyGlnAsn-----GlnTyArgAspAlaArgAsnGln 1011
QY 994 ACGGGAACCT----- 1002
Db 1011 uTyrgluAspGlnTyrglyGlnGlnSerGlnTyrglyGlnAspGlnTyrglyGlnAspGln 1031
QY 1003 -----GACTGTGCTCCAGGAGAGAAGCT----- 1026
Db 1031 nTyrgluAspGlnTyrglyGlnAlaGlyProGlyGlnAspGlnTyrglyGlnAspArgProAr 1051
QY 1027 -----GGAGTTCTGCAGCACAGCGGTGCTGCTGAGAGATGAGCGGACGGA 1082
Db 1051 gPheGlyGlnAspAlaProGlnAsnAspGlnPheAlaArgProGluProAlaAlaProGln 1071
QY 1083 GGG-----CTACATTCCAAGCAACATCTCGA-----GCCCTACACG 1121
Db 1071 nGlnAspGlyGlyAlaPheValArgSerAspValPheGlyGlnProAlaProAsnGln 1091
QY 1122 CGGGGACCCCTGGGACCCAGGCGCCAGTCAACCTCTCGGTTCCTCAATGCTTCGACTAGCT 1181
```

```

Db 1091 yGlnAlaProAlaAspArgPheAlaAlaSerGlnGlyTyrAspAsnGlySerThr 1109
QY 1182 CGAGGCTGAAGAGGTGCACAGCTGGCTGCAGCGAGAGAACTTCCTCAGCTCCACGGTGA 1241
Db 1110 -----GlyGlnHisGlyLeuProGlyArgGln----- 1118
QY 1242 GGACACTGGTCCCTGAGCGGGAGCAGCTACTTCGCGATAAG----- 1284
Db 1119 -----AspProAlaValThrGlyGlnPheGluArgProGlnAl 1131
QY 1285 -----ACCTGGGAGCTACAGATGCTATGTCACAGAGAGGCCACGAATCCTGCTCC 1337
Db 1131 aGlyGlyThrGlnGlyGlyAsnAsp-----PheGlyAlaProArgProAlaPr 1148
QY 1338 GGCTGGAGGCTGTCACAAGAGCTCTGGGATAGCCCTTAGGCACACAGCTTACACACCTC 1397
Db 1148 o-----GlnArgProValProGlnGluProAlaGlyGlnAsnThrAsnGlyTh 1164
QY 1398 CAAGAACCCAGGCCCG-----CTGATGCAAGATGG 1427
Db 1164 rProAlaGlnArgProGlyAspGlyTrpAlaLeuProProAlaSerGlyProGlyAspGl 1184
QY 1428 CAGATCT-----GATACC-----CATTAGAGCCCGAGAAATTC 1460
Db 1184 yArgThrProLeuTyrAspThrLeuGluThrAsnTrpPheHisGlyAspArgGluSerAr 1204
QY 1461 CTCTCTCGATCCACTTTCAGCAACCCACACCCAGCTCAGCTCAGCAGCAAAACAATG 1520
Db 1204 gAlaProGlnAlaAlaProAlaAlaSerProGluProGlnAlaAlaGlnProGlnThrPr 1224
QY 1521 GACAGGCCAGAGGCTGAAGCAAAACAGT 1548
Db 1224 oAlaAlaProGlnArgProAlaThrSer 1233

RESULT 15
T19140
hypothetical protein C09G5.6 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T19140
R:Palmer, S.
submitted to the EMBL Data Library, November 1994
A:Reference number: Z19080
A:Accession: T19140
A:Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-963 <WIL>
A:Cross-references: EMBL:Z46791; PIDN:CAA86755.1; GSPDB:GN00020; CESP:C09G5.6
A:Experimental source: clone C09G5
C:Genetics:
A:Gene: CESP-C09G5.6
A:Map position: 2
A:Introns: 48/3; 862/3; 898/1

Alignment Scores:
Pred. No.: 3-58e-05 Length: 963
Score: 185.00 Matches: 141
Percent Similarity: 30.94% Conservative: 44
Best Local Similarity: 23.58% Mismatches: 206
Query Match: 5.92% Indels: 208
DB: 2 Gaps: 31

US-09-762-021A-1 (1-1710) x T19140 (1-963)
QY 34 TCTGGAGGAGAGCT-----GGAGCAAAAGACTTCAGCTTTGGAGGCT----- 75
Db 154 SerGlyPheArgAlaProProAlaAlaThrSerThrTyrArgProProHisGlySer 173
QY 76 -----TCAGCCAGG-----CCAGACAG 93
Db 174 AsnTyrAspAsnTyrGlyArgGluProAlaSerSerArgArgProTyrProProGlnGln 193
```

```

QY 94 ATGAGGGGGCGCTGCTATGGAAGGCCGCTCCTATGGAGCAGGACGCTATCTGTAGGCC 153
Db 194 ProProSerThrSerAlaProHisSerSerProAsnAsnArgThrSerLeuTyrAsnPro 213
QY 154 GGGGATCCTCCAGAACACCCACCAGAGAGACCCCTAGAGCA-----CAG 198
Db 214 GlnProProProGlyThrGlyTyrProThrAsnProArgValProTyrAsnProProGln 233
QY 199 CCT-----CCCAACCATCC-----AAGCCCTCCACCGCCACAC 234
Db 234 ProAsnTyrThrArgGlnProThrTyrProGluAspAsnArgAlaProTyrLysProThr 253
QY 235 CAGTCCCGAGAACCAAGTCCTTACTCTGCTCCTCCTCAAGCGCGTCTCTTCCCCCA 294
Db 254 ArgSerProAsnThrPro-----ProArg 262
QY 295 GCACCCAGAGAGGAGGAGGAAGTCTGAACCATGCTCCTAAGGACATTGAGCTGTCAT 354
Db 263 GlnProSerGlyGlyTyrAspSerAspGlyGlnThrProProSerSerProArgIleTyr 282
QY 355 GGGAAAGCTGGAGAGGCCCGCCAGCAAGACACGAGGAAGAAATTTGGGAAAAAAA 414
Db 283 -----AsnThr 284
QY 415 CAAGGACACGAGGAGTCTCACCCAGGCACAGTACATGCTCTCCAGAGATCAAGTA 474
Db 285 ArgArgProAsnAsnHisGlyProGly-----TyrProGluAspGlnVal 299
QY 475 CAGCTTCAACCTCCCTGGGAGGCTGGCCACCTGGCTGAAGGAGACAAAGTGCCTGAGCT 534
Db 300 ProThrAlaPro-----ValProGlyGlnGln----- 309
QY 535 CGTACACATCCTCTCAAGTCCCTGAACCTTCATCTCGCCAGGTGCCTGAGGCTGGCT 594
Db 310 ArgValProProThrGlnThrArgAsn---ProProAsnProThrAsnThrArgGlnPro 328
QY 595 AGCAGCCCAAGTGATCTCACCCCTCCTCACCCCTAAAGCTATCAACCTCTCAGCTCTG 654
Db 329 SerArgPro-----ValProProThrSerAspGlyHisIleGluAlaThrPro 345
QY 655 TCTAAGCCCACTGAGATGAACCTTT-----CGATGGGTTGG 692
Db 346 TyrAsnProSerAlaGlnTyrProThrGlyLysArgGlySerHisProGlyPheGlyPro 365
QY 693 GCCCAGCTGGACCACTAGCCGGCGGCGACTGGCAGGCGCATGAGCCCTGCCCTACCAAC 752
Db 366 GlnArgProArgProGlyThrArgProArgProArgGlyAsnProCysAspGlnCysSerAlaGln 385
QY 753 CCACATTTCTCAGATGACTGGCAACTTCCAGAGCCCTCCAGCCCAAGCACCCTTAGGATACC 812
Db 386 ProAsnHisCysProSerGly-----ProProGlyProArgGly-ArgProGl 401
QY 813 AGGACCTGTTTCCCTTCCGCGGGAAGTCAATAGTTAGGAGGACCTCACATTTCTTC 872
Db 401 yProProGlyPhePro-----GlyGlnAspGlyProArgGlyLeuArgGlyLeuAsnGl 419
QY 873 AGGAGAAGACACAAACATGACCTCAGCCTGGGAGCCCAACCTCAGGCGCTCCAGCC 932
Db 419 yGlyTyrSerGlyValGln---ProSerSerTyrAspProValIleGlyCysValGln--- 437
QY 933 CCAAACTGCCCGCAGCCCTGAAANTTCAAGTCTTGTACGAGTTTGAAGCTAGGAACC 992
Db 438 -----CysProIleGlyProGlyGluArg-----GlyPr 448
QY 993 CAGGGAACTGACTGTGTCAGGAGGAGAGAACTGGAGTTCTGGACCACAGCAAGCGGT 1052
Db 448 oAspGlyThrProGlyValProGly---GluAspGlyIleAspGlyGluGlnGly----- 465
QY 1053 GTGTGCTGTGAAGATAGGCGGAGCGGCGGTACATTCCAGCAACATCTCTGGAGC 1112
Db 466 -ValAsnGlyGlnAsp-----GlyGlnProGlyAl 475
QY 1113 CCTACAGCC-----GGGAGCCCTGGGAGCCAGGCCAGTCAC 1151
```

```
Db 475 aProGlyAlaProGlyTyrHisGlyMetAsnGlySerProGlyThrProGlyLysProG1 495
QY 1152 CCTCTCGGGTTCCAACTTTCGACTTAGCTCGAGGCCCTGAAGAGGTCACAGACTGGCTGC 1211
Db 495 yLeuProGlyArgAsnGlyGlnSer----- 503
QY 1212 AGCAGAGAACTTCTCCACTGC----- 1233
Db 504 -----CysLysSerIleProGlyProProGlyGlnProGlyValMe 517
QY 1234 -----CACGGTAGGAC----- 1245
Db 517 tGlyValProGlyArgAspGlyThrAspGlyGluHisGlyGlnAspGlyse 537
QY 1246 -----ACTTGGTCCCTGACGGGAGCCAGC 1271
Db 537 rProGlyIleGlnGlyProProGlyArgAspGlyThrSerGlyProAspGlyGlnProG1 557
QY 1272 TACTTCGCATAAGACCTGGGAGCTACAGATGCTATGTCACAGAGGCCCCACGAATCC 1331
Db 557 yValSer-----AlaProGlyAlaProGlyThrAspGlyGlyTyrCysPr 572
QY 1332 TGTCCTCCGGCTGGAGCTGTCAAGAGGATGCTGGGGAT----- 1368
Db 572 oCysProLysArgSerSerLysPheAspPheAsnAspAlaTyrAsnAspAspGluLysAr 592
QY 1369 -----AAGCCCTTAGGCACACAGCTTAGACACCTCCCAAGAACCCAGGCCCC 1412
Db 592 gGlyLeuGluGluHisArgProArgGlyTyrAspSerGluArgAlaGluGluProArgPr 612
QY 1413 GCTGATCAAGATGGCAGATCTGATACCCATTAGAGCCCGAGAAATTCCTCTTCTGGATC 1472
Db 612 o---ArgGlnThrValArgThrAsnThrTyr-----AspGluAsnSerGlyAlaGluHi 629
QY 1473 CCAGTTTTCAGCAAC-----CCACACCCAGCTCACAGCAAAACAA 1518
Db 629 sGlnArgArgProAsnTyrGluProSerAlaGluValAlaProProArgGln 646
```

Search completed: February 25, 2003, 19:44:33
Job time : 61.5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame_plus_n2p model

Run on: February 25, 2003, 19:43:04 ; Search time 32 Seconds

(without alignments)
3320.587 Million cell updates/sec

Title: US-09-762-021A-1

Perfect score: 3127

Sequence: 1 ggcagagcactgaagacca.....aaaaaaaaaataangataaa 1710

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 156504 seqs, 31069816 residues

Total number of hits satisfying chosen parameters: 313008

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+n2p.model -DEV=xlh
-O=/cgn2_1/USPTO_spool/US09762021/runat_24022003_153040_7696/app_query.fasta_1.1863
-DB=PublishedApplications_AA -QFM=fastan -SUFFIX=rapb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=2000000000 -USER=US09762021@cgn_1.1.9 @runat_24022003_153040_7696
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=10 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

PublishedApplications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB pep.*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB pep.*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB pep.*
4: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB pep.*
5: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB pep.*
6: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB pep.*
7: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB pep.*
8: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
9: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
10: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB pep.*
11: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
12: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB pep.*
13: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB pep.*
14: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2229	71.3	563	10	US-09-998-598-2591
2	912	29.2	179	9	US-09-764-868-757
3	602	19.3	121	9	US-09-764-868-756
4	598	19.1	134	9	US-09-764-868-1135

5	256.5	8.2	280	9	US-09-764-868-1125	Sequence 1125, Ap
6	256.5	8.2	283	9	US-09-764-868-712	Sequence 712, App
7	186.5	5.2	1023	9	US-09-893-519A-14	Sequence 14, Appl
8	171.5	5.2	1274	9	US-10-020-215-2	Sequence 2, Appli
9	164	5.2	2005	10	US-09-735-367B-3	Sequence 3, Appli
10	163	5.2	2063	10	US-09-735-367B-2	Sequence 2, Appli
11	161.5	5.2	846	10	US-09-858-664A-3	Sequence 3, Appli
12	155.5	5.0	1690	10	US-09-788-043C-5	Sequence 5, Appli
13	154.5	4.9	1806	10	US-09-919-497-56	Sequence 56, Appl
14	154	4.9	1070	10	US-09-735-367B-6	Sequence 6, Appli
15	153.5	4.9	941	12	US-10-124-557-14	Sequence 14, Appl
16	153.5	4.9	1022	12	US-10-124-557-84	Sequence 84, Appl
17	153.5	4.9	1038	12	US-10-124-557-74	Sequence 74, Appl
18	153.5	4.9	1049	12	US-10-124-557-58	Sequence 58, Appl
19	153.5	4.9	1140	12	US-10-124-557-104	Sequence 104, App
20	153.5	4.9	1270	12	US-10-124-557-44	Sequence 44, Appl
21	153.5	4.9	1311	12	US-10-124-557-42	Sequence 42, Appl
22	153.5	4.9	1313	12	US-10-124-557-142	Sequence 142, App
23	153.5	4.9	1314	12	US-10-124-557-50	Sequence 50, Appl
24	153.5	4.9	1320	12	US-10-124-557-46	Sequence 46, Appl
25	153.5	4.9	1320	12	US-10-124-557-60	Sequence 60, Appl
26	153.5	4.9	1354	12	US-10-124-557-48	Sequence 48, Appl
27	153.5	4.9	1361	12	US-10-124-557-40	Sequence 40, Appl
28	153.5	4.9	1363	12	US-10-124-557-52	Sequence 52, Appl
29	153.5	4.9	1404	12	US-10-124-557-2	Sequence 2, Appli
30	153.5	4.9	1404	12	US-10-124-557-62	Sequence 62, Appl
31	152.5	4.9	881	10	US-09-816-860A-2	Sequence 2, Appli
32	151	4.8	661	9	US-10-160-865-14	Sequence 14, Appl
33	151	4.8	1907	9	US-09-832-292-39	Sequence 39, Appl
34	150.5	4.8	827	9	US-10-171-384-3	Sequence 3, Appli
35	149	4.8	419	9	US-10-002-344A-237	Sequence 237, App
36	147	4.7	529	10	US-09-861-597-2	Sequence 2, Appli
37	146.5	4.7	775	9	US-09-738-626-3773	Sequence 3773, Ap
38	146	4.7	575	9	US-10-028-072-128	Sequence 128, App
39	146	4.7	575	9	US-10-121-049-128	Sequence 128, App
40	146	4.7	575	9	US-10-123-904-128	Sequence 128, App
41	146	4.7	575	9	US-10-140-470-128	Sequence 128, App
42	146	4.7	575	9	US-10-175-746-128	Sequence 128, App
43	146	4.7	575	9	US-10-176-918-128	Sequence 128, App
44	146	4.7	575	9	US-10-176-921-128	Sequence 128, App
45	146	4.7	575	9	US-10-137-865-128	Sequence 128, App

ALIGNMENTS

RESULT 1
US-09-998-598-2591
; Sequence 2591, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2591
; LENGTH: 563
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-998-598-2591

Alignment Scores:
Pred. No.: 5,11e-153 Length: 563
Score: 2229.00 Matches: 428
Percent Similarity: 93.45% Conservative: 0
Best Local Similarity: 93.45% Mismatches: 0
Query Match: 71.28% Indels: 30

```
DB: 10 Gaps: 1
US-09-762-021a-1 (1-1710) x US-09-998-598-2591 (1-563)
QY 2 CGAGGCGACTGAAGACCAGCTCGAAGAGCTCTGAGGAAGAGCTGGAGCAAGACCT 61
Db 136 AlaGluArgLeuLysThrSerLeuGlnLysAlaLeuGluGluGluGlnArgPro 155
QY 62 CGACTTGAGGCGCTTACGCCAGGCAGACAGATGGAGGGCGCTGTATGGAAAGCCG 121
Db 156 ArgLeuGlyGlyLeuGlnProGlyGlnAspArgTrpArgGlyProAlaMetGluArgPro 175
QY 122 CTCCTATGGAGCAGGACCGCTATCTGGAGCCGGGGATCCCTCCAGAACGCCACCAG 181
Db 176 LeuProMetGluGlnAlaArgTyrLeuGluProGlyTyrLeuGluGlnProHisGln 195
QY 182 AGGACCTTAGAGCACAGCTCCACCATCCCAAGGGCGCTCTTCCCGGAGGACCCA 241
Db 196 ArgThrLeuGluHisSerLeuProProSerProArgProLeuProArgHisThrSerAla 215
QY 242 CGAGAACCAAGTGCCTTACTCTCCCTCCAAGGGCGCTCTTCCCGGAGGACCCA 301
Db 216 ArgGluProSerAlaPheThrLeuProProProArgArgSerSerProGluAspPro 235
QY 302 GAGAGGACGAGGAAGTGTCTGAACCATGTCTTAAGGACATTTGAGCTGTTCATGGGAAAG 361
Db 236 GluArgAspGluGluValLeuAsnHisValLeuArgAspIleGluLeuPheMetGlyLys 255
QY 362 CTGAGAGGCGCCAGGCAAGACCAGGAGGAGAAATTTGGGAAAAAACAAGAC 421
Db 256 LeuGluLysAlaGlnAlaLysThrSerArgLysLysPheGlyLysLysAsnLysAsp 275
QY 422 CAGGAGGTCTCACCCAGGACAGTACATTGACTGCTTCCAGAAAGATCAAGTACAGCTC 481
Db 276 GlnGlyGlyLeuThrGlnAlaGlnTyrIleAspCysPheGlnLysIleLysTyrSerPhe 295
QY 482 AACCTCTGGGAGGCTGGCCACCTGGCTGAAGAGACAAGTGGCCCTGAGCTCGTACAC 541
Db 296 AsnLeuLeuGlyArgLeuAlaThrTrpLeuLysGluThrSerAlaProGluLeuValHis 315
QY 542 ATCTCTTCAAGTCCCTGAAGTTCATCTGGCCAGGTGCGCTGAGGCTGGCTAGCAGCC 601
Db 316 IleLeuPheLysSerLeuAsnPheIleLeuAlaArgCysProGluAlaGlyLeuAlaAla 335
QY 602 CAAGTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGTACAGTCTGCTGAAGC 661
Db 336 GlnValIleSerProLeuLeuThrProLysAlaIleAsnLeuLeuLeuSerCysLeuSer 355
QY 662 CCACCTGAGAGTAACCTTTGGATGGGTGGGGCCAGCCTGGACCACTAGCCGGCGCGAC 721
Db 356 ProProGluSerAsnLeuTrpMetGlyLeuGlyProAlaTrpThrThrSerArgAlaAsp 375
QY 722 TGGACAGCGATAGCCCGCTCGCTACCAACCCACATTCAGATGACTGCAACTTCCA 781
Db 376 TrpThrGlyAspGluProLeuProTyrGlnProThrPheSerAspAspTrpGlnLeuPro 395
QY 782 GAGCCCTCCAGCCCAAGCACCTTAGATACAGGACCTGTTCCTCCCTCGCGGGGAAGT 841
Db 396 GluProSerSerGlnAlaProLeuGlyTyrGlnAspProValSerLeu----- 411
QY 842 CATAGTTAGGAGCACCTTCACACTTTCCTCAGGAGAAGACACAAACCATGACCCCTCAG 901
Db 411 ----- 411
QY 902 CCTGGGACCCCAACTCGAGGCCCTCGAGCCCAAAACCTTGGCCAGCCAGCCCTGAAATG 961
Db 412 -----ArgProSerSerProLysProAlaGlnProAlaLeuLysMet 425
QY 962 CAAGTCTTGTAAGACTTAGGAACCCAGCGGAACCTGACTGTGGTCCAGGGAGAG 1021
Db 426 GlnValLeuTyrGluPheGluAlaArgAsnProArgGluLeuThrValValGlnGlyGlu 445
QY 1022 AAGCTGGAGGTTCTGGACCACAGCAAGCGGTGGTGGTGTGAAGATGAGCGGGGACGG 1081
```

```
Db 446 LysLeuGluValLeuAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArg 465
QY 1082 AGCGGCTTACATTCACAGCAACATCCTGGAGCCCTACAGCGGGGACCCCTGGGACCAG 1141
Db 466 SerGlyTyrIleProSerAsnIleLeuGluProLeuGlnProGlyThrProGlyThrGln 485
QY 1142 GGCAGTCCACCCCTCTCGGGTTCCAATGCTTCGACTTAGCTCGAGGCGCTGAAGAGTCA 1201
Db 486 GlyGlnSerProSerArgValProMetLeuArgLeuSerSerArgProGluGluValThr 505
QY 1202 GACTGGCTGAGGAGAGAACTTCTCCACTGCCAGGTGAGGACACTTGGGTCCCTGACG 1261
Db 506 AspTrpLeuGlnAlaGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThr 525
QY 1262 GGAGCGCAGCTACTTCGCATAAGACCTGGGAGCTGCAGAGGATGCTGCTCCACAGGAGCC 1321
Db 526 GlySerGlnLeuLeuArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAla 545
QY 1322 CCAGAACTCTGCCCGCTGGAGGCTGTGCAGAGGATGCTGGGGATAAGCCCT 1375
Db 546 ProArgIleLeuSerArgLeuGluAlaValArgArgMetLeuGlyIleSerPro 563
RESULT 2
US-09-764-868-757
; Sequence 757, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 757
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-757
Alignment Scores:
Pred. No.: 3,07e-58 Length: 179
Score: 912.00 Matches: 179
Percent Similarity: 92.27% Conservative: 0
Best Local Similarity: 92.27% Mismatches: 0
Query Match: 29.17% Indels: 15
DB: 9 Gaps: 1
US-09-762-021A-1 (1-1710) x US-09-764-868-757 (1-179)
QY 796 AGACACCTTAGGATACAGGACCTGTTCCCTTCGCGGGGAGTGCATAGTTAGGGAG 855
Db 1 SerThrLeuArgIleProGlyProCysPhePro----- 11
QY 856 CACCTCACACTTTCCTCAGGAGAGACACACACCATGACCTCAGCCTGGGACCCCAA 915
Db 12 -----Ser-GluLysThrHisAsnHisAspProGlnProGlyAspProAs 26
QY 916 CTCAGGCGCTCCAGCCCCAAACCTGCCAGCCAGCCCTGAAAATGCAAGTCTTGTACGA 975
Db 26 nSerArgProSerSerProLysProAlaGlnProAlaLeuLysMetGlnValLeuTyrG 46
QY 976 GTTTGAAGCTTAGNAACCCAGCGGAACCTGCTGCTCCAGGAGAGAGCTGGAGGTTCT 1035
Db 46 uPheGluAlaArgAsnProArgGluLeuThrValValGlnGlyLysLeuGluValLe 66
QY 1036 GGACCACAGCAAGCGGTGGTGGTGAAGATGAGCGGGGAGCGGCGGCTACATTCC 1095
Db 66 uAspHisSerLysArgTrpTrpLeuValLysAsnGluAlaGlyArgSerGlyTyrIlePr 86
QY 1096 AAGCAACATCCTGGAGCCCTACAGCGGGGACCCCTTGGGACCCAGGCGGACCCCTC 1155
```

```
|||||
Db 86 oserAsnIleLeuGluProLeuGlnProGlyThrProGlyThrGlnGlyGlnSerProSe 106
QY 1156 TCGGGTTCCAACTTCGACTTAGCTCGAGCGCTGAAGAGGTGCACACACTGGCTGCAGGC 1215
Db 106 rArgValProMetLeuArgLeuSerArgProGluGluValThrAspTrpLeuGlnAl 126
QY 1216 AGAAACTTCTCCACTGCCACGGTGAGCACACTTGGGTCCCTGACGGGAGCCAGCTACT 1275
Db 126 aGluAsnPheSerThrAlaThrValArgThrLeuGlySerLeuThrGlySerGlnLeuLe 146
QY 1276 TCGCATTAAGACCTGGGAGGTACAGATGCTATGTCCACAGAGGCCCCACGAATCCGTGC 1335
Db 146 uArgIleArgProGlyGluLeuGlnMetLeuCysProGlnGluAlaProArgIleLeuSe 166
QY 1336 CCGCTCGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCT 1375
Db 166 rArgLeuGluAlaValArgMetLeuGlyIleSerPro 179
RESULT 3
US-09-764-868-756
; Sequence 756, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 756
; LENGTH: 121
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (4)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-756
Alignment Scores:
Pred. No.: 6,09e-36 Length: 121
Score: 602.00 Matches: 116
Percent Similarity: 98.33% Conservative: 2
Best Local Similarity: 96.67% Mismatches: 2
Query Match: 19.25% Indels: 0
DB: 9 Gaps: 0
US-09-762-021A-1 (1-1710) x US-09-764-868-756 (1-121)
QY 1016 GGAGAGAAGCTGGAGGTTCTGGACCAGCAAGCGGTGGTGGTGAAGATGAGCGC 1075
Db 2 GlyAsp***TrpGlnValLeuAspHisSerLysArgTrpTrpLeuValLysAsnGluAla 21
QY 1076 GGACGGAGCGGTACATTCCTCAAGCAACATCTGTGAGGCCCTACAGCGGGGCCCTGGG 1135
Db 22 GlyArgSerGlyTyrlleProSerAsnIleLeuGluProLeuGlnProGlyThrProGly 41
QY 1136 ACCAGGCCAGTCACCTCTCGGGTTCCAAATGCTTGGACTTACCTGAGCCCTGAAGAG 1195
Db 42 ThrGlnGlyGlnSerProSerArgValProMetLeuArgLeuSerSerArgProGluGlu 61
QY 1196 GTCACACACTGGCTGCAGGCAGAGAAGTCTCCACTGCCAGGTGAGGACACTTGGGTCC 1255
Db 62 ValThrAspTrpLeuGlnAlaGluAsnPheSerThrAlaThrValArgThrLeuGlySer 81
QY 1256 CTGACGGGGAGCCAGCTACTTCGCATAAGACCTGGGGAGCTACAGATGCTATCTCCACAG 1315
Db 82 LeuThrGlySerGlnLeuLeuArgIleArgProGlyGluLeuGlnMetLeuCysProGln 101
QY 1316 GAGGCCCCACGAATCTGTCTCCGGCTGGAGGCTGTCAAGAGGATGCTGGGGATAAGCCCT 1375
```

```
|||||
Db 102 GluAlaProArgIleLeuSerArgLeuGluAlaValArgMetLeuGlyIleSerPro 121
RESULT 4
US-09-764-868-1135
; Sequence 1135, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1135
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (84)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
; NAME/KEY: SITE
; LOCATION: (128)
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-764-868-1135
Alignment Scores:
Pred. No.: 1,21e-35 Length: 134
Score: 598.00 Matches: 111
Percent Similarity: 99.11% Conservative: 0
Best Local Similarity: 99.11% Mismatches: 1
Query Match: 19.12% Indels: 0
DB: 9 Gaps: 0
US-09-762-021A-1 (1-1710) x US-09-764-868-1135 (1-134)
QY 827 CTTCCGGCGGGAAGTCATAGTTAGGAGGACCTCACACTTCTCAGGAGAAGACACAC 886
Db 14 LeuArgArgGlySerHisArgLeuGlySerThrSerHisPheProGlnGluLysThrHis 33
QY 887 AACCATGACCTCAGCCTGGGACCCCACTCCAGGCCCTCCAGGCCCAACCTGCCAG 946
Db 34 AsnHisAspProGlnProGlyAspProAsnSerArgProSerSerProLysProAlaGln 53
QY 947 CCAGCCCTGAAAATGCAAGTCTTTGTACGAGTTTGAAGCTAGGAACCCACCGGAAGTGA 1006
Db 54 ProAlaLeuLysMetGlnValLeuTyrGluPheGluAlaArgAsnProArgGluLeuThr 73
QY 1007 GTGCTCCAGGAGAGAGCTGGAGGTTCTGGACCACAGAGCGGTGGTGGTGAAG 1066
Db 74 ValValGlnGlyLysLeuGluValLeu***HisSerLysArgTrpTrpLeuValLys 93
QY 1067 AATGAGCGGGAGCGGAGCGGCTACATTCCAAAGCAACATCTCGAGGCCCTACAGCCGGG 1126
Db 94 AsnGluAlaGlyArgSerGlyTyrlleProSerAsnIleLeuGluProLeuGlnProGly 113
QY 1127 ACCCTCTGGGACCCAGGCCAGTCACTCCCTCTCGGGTT 1162
Db 114 ThrProGlyThrGlnGlyGlnSerProSerArgVal 125
RESULT 5
US-09-764-868-1125
; Sequence 1125, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
```

; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 1125
; LENGTH: 280
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-1125

Alignment Scores:
Pred. No.: 5.81e-11 Length: 280
Score: 256.50 Matches: 76
Percent Similarity: 43.90% Conservative: 32
Best Local Similarity: 30.89% Mismatches: 67
Query Match: 8.20% Indels: 71
DB: 9 Gaps: 10

US-09-762-021A-1 (1-1710) x US-09-764-868-1125 (1-280)

```
Qy 818 CCTGTTTCCCTTCGGCGGGAAGTCATAGTTAGGGAGCACCTCACACACTTTCTCCTCAGGAG 877
      |||||
Db 15 ProValSer-----ArgGlnSerIleArg---AsnSerGlnLysHisSerPro----- 29

Qy 878 AAGACACACACCAATGACCTCAGCCCTGGGAC-----CCCAACTCAGGCCCTCC 928
      |||||
Db 30 ---ThrSerGluProThrPro---ProGlyAspAlaLeuProProValSerSerProHis 47

Qy 929 AGCCCCAAACCTGCCAGCCA-----GCCCTGAAATGCAAGTCTTCTACGAG 976
      |||||
Db 48 ThrHisArgGlyTyrglnProThrProAlaMetAlaLysTyrvAllyIleLeutyAsp 67

Qy 977 TTTGAAGCTAGGAACCCAGCGGAAGTCTGCTGTCAGGAGAGAGCTGGAGTTCTG 1036
      |||||
Db 68 PheThrAlaArgAsnAlaAsnGluLeuSerValLeuLysaspGluValLeuLysValLeu 87

Qy 1037 GACCACACCAAGCGTGTGCTGTGAAGAAATGAGCGGGAGCGGCTACATTCCA 1096
      |||||
Db 88 GluAspGlyArgGlnTrpTrpLysLeuArgSerArgSerGlyGlnAlaGlyTyrvAlPro 107

Qy 1097 AGCAACATCTCTG----- 1108
      |||||
Db 108 CysAsnIleLeuGlyGluAlaArgProGluAspAlaGlyAlaProPheGluGlnAlaGly 127

Qy 1109 -----GAGCCCTACAGCCGGGAGCCCTGGGACC 1138
      |||||
Db 128 GlnLysTyrrpGlyProAlaSerProThrHisLysLeuProProSerPheProGlyAsn 147

Qy 1139 CAGGCCAG----- 1147
      |||||
Db 148 LysAspGluLeuMetGlnHisMetAspGluValAsnAspGluLeuIleArgLysIleSer 167

Qy 1148 -----TCACCTCTCTCG-----GTTCCAATG----- 1168
      |||||
Db 168 AsnIleArgAlaGlnProGlnArgHisPheArgValGluArgSerGlnProValSerGln 187

Qy 1169 ---CTTCGACTTAGCTGAGGCCTGAAGAGTGCACAGACTGCTGCAGGCAGAGAACTTC 1225
      |||||
Db 188 ProLeuThrTyrgluSerGlyProAspGluValArgAlaTrpLeuGluAlaLysAlaPhe 207

Qy 1226 TCCACTGCCAGGTGAGACACTTGGTCTCCTGAGGGGAGCCAGCTACTTCGATAGA 1285
      |||||
Db 208 SerProArgIleValGluAsnLeuGlyIleLeuThrGlyProGlnLeuPheSerLeuAsn 227

Qy 1286 CTGGGGAGCTACAGATCTATGTCCACGAGGCCCCACCAATCTCTCCCGCTGGAG 1345
      |||||
Db 228 LysGluGluLeuLysValCysGlyGluGluGlyValArgValTyrrSerGlnLeuThr 247

Qy 1346 GCTGTGAGAAGGATGCTG 1363
      |||||
Db 248 MetGlnLysAlaPheLeu 253

RESULT 6
US-09-764-868-712
```

; Sequence 712, Application US/09764868
; Patent No. US20020168711A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT232
; CURRENT APPLICATION NUMBER: US/09/764,868
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - refer to PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1510
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 712
; LENGTH: 283
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-868-712

Alignment Scores:
Pred. No.: 5.82e-11 Length: 283
Score: 256.50 Matches: 76
Percent Similarity: 43.90% Conservative: 32
Best Local Similarity: 30.89% Mismatches: 67
Query Match: 8.20% Indels: 71
DB: 9 Gaps: 10

US-09-762-021A-1 (1-1710) x US-09-764-868-712 (1-283)

```
Qy 818 CCTGTTTCCCTTCGGCGGGAAGTCATAGTTAGGGAGCACCTCACACACTTTCTCCTCAGGAG 877
      |||||
Db 18 ProValSer-----ArgGlnSerIleArg---AsnSerGlnLysHisSerPro----- 32

Qy 878 AAGACACACACCAATGACCTCAGCCCTGGGAC-----CCCAACTCAGGCCCTCC 928
      |||||
Db 33 ---ThrSerGluProThrPro---ProGlyAspAlaLeuProProValSerSerProHis 50

Qy 929 AGCCCCAAACCTGCCAGCCA-----GCCCTGAAATGCAAGTCTTGTACGAG 976
      |||||
Db 51 ThrHisArgGlyTyrglnProThrProAlaMetAlaLysTyrvAllyIleLeutyAsp 70

Qy 977 TTTGAAGCTAGGAACCCAGCGGAAGTCTGCTGTCAGGAGAGAGCTGGAGTTCTG 1036
      |||||
Db 71 PheThrAlaArgAsnAlaAsnGluLeuSerValLeuLysaspGluValLeuLysValLeu 90

Qy 1037 GACCACACCAAGCGTGTGCTGTGAAGAAATGAGCGGGAGCGGCTACATTCCA 1096
      |||||
Db 91 GluAspGlyArgGlnTrpTrpLysLeuArgSerArgSerGlyGlnAlaGlyTyrvAlPro 110

Qy 1097 AGCAACATCTCTG----- 1108
      |||||
Db 111 CysAsnIleLeuGlyGluAlaArgProGluAspAlaGlyAlaProPheGluGlnAlaGly 130

Qy 1109 -----GAGCCCTACAGCCGGGAGCCCTGGGACC 1138
      |||||
Db 131 GlnLysTyrrpGlyProAlaSerProThrHisLysLeuProProSerPheProGlyAsn 150

Qy 1139 CAGGCCAG----- 1147
      |||||
Db 151 LysAspGluLeuMetGlnHisMetAspGluValAsnAspGluLeuIleArgLysIleSer 170

Qy 1148 -----TCACCTCTCTCG-----GTTCCAATG----- 1168
      |||||
Db 171 AsnIleArgAlaGlnProGlnArgHisPheArgValGluArgSerGlnProValSerGln 190

Qy 1169 ---CTTCGACTTAGCTGAGGCCTGAAGAGTGCACAGACTGCTGCAGGCAGAGAACTTC 1225
      |||||
Db 191 ProLeuThrTyrgluSerGlyProAspGluValArgAlaTrpLeuGluAlaLysAlaPhe 210

Qy 1226 TCCACTGCCAGGTGAGACACTTGGTCTCCTGAGGGGAGCCAGCTACTTCGATAGA 1285
      |||||
Db 211 SerProArgIleValGluAsnLeuGlyIleLeuThrGlyProGlnLeuPheSerLeuAsn 230

Qy 1286 CTGGGGAGCTACAGATCTATGTCCACGAGGCCCCACCAATCTCTCTCCCGCTGGAG 1345
      |||||
```

Db 231 LysGluGluLeuLysLysValCysGlyGluGluGlyValArgValTyrSerGlnLeuThr 250
QY 1346 GCTGTCTCAGAGGATGCTG 1363
Db 251 MetGlnLysAlaPheLeu 256
RESULT 7
US-09-893-519A-14
; Sequence 14, Application US/09893519A
; Publication No. US20030027243A1
; GENERAL INFORMATION:
; APPLICANT: ANADYS PHARMACEUTICALS, INC.
; APPLICANT: THOMPSON, Craig
; APPLICANT: MOORE, Jeffrey
; APPLICANT: BUURMAN, Ed T.
; APPLICANT: BRADLEY, John
; APPLICANT: DESILVA, Thamara
; APPLICANT: HARRIS, Sandra
; APPLICANT: KOMARNITSKY, Svetlana
; APPLICANT: MENDILLO, Marc
; APPLICANT: MOORE, Daniel
; APPLICANT: MCCOY, Melissa
; APPLICANT: SANDERSON, Karen
; APPLICANT: HAQ, Tariq
; APPLICANT: ZHU, Shuhao
; APPLICANT: LONG, Fan
; APPLICANT: DAVIDOV, Eugene
; TITLE OF INVENTION: ANTIFUNGAL COMPOUNDS AND METHODS OF USE
; FILE REFERENCE: 0342/1G548-US2
; CURRENT APPLICATION NUMBER: US/09/893,519A
; CURRENT FILING DATE: 2001-06-28
; PRIOR APPLICATION NUMBER: US 60/215,164
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: US 60/224,457
; PRIOR FILING DATE: 2000-08-10
; NUMBER OF SEQ ID NOS: 146
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 14
; LENGTH: 1023
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Corresponds to SEQ ID NO: 87
; PUBLICATION INFORMATION:
; DATABASE ACCESSION NUMBER: Human Genbank/CAA72189
; DATABASE ENTRY DATE: 1997-06-25
; RELEVANT RESIDUES: (1)..(1023)
US-09-893-519A-14
Alignment Scores:
Pred. No.: 8 58e-06 Length: 1023
Score: 186.50 Matches: 136
Percent Similarity: 33.87% Conservative: 54
Best Local Similarity: 24.24% Mismatches: 224
Query Match: 5.96% Indels: 148
DB: 9 Gaps: 27
US-09-762-021a-1 (1-1710) x US-09-893-519A-14 (1-1023)
QY 22 CTCTCAGAGAGCTCTGGAGGAGAGCTGGAGCAAGACCTCGACTTTGGAGGCGCTTCAGCG 81
Db 78 ProAlaGluGlyAlaProGlyAlaAla-----ProGluPro 89
QY 82 AGGCCAGGACAGATGAGGGGCGCTCTATGGAAAGCGCGCTCCCTATGAGCAGCAGC 141
Db 90 ProProAla-----GlyArgAlaArgPro----- 97
QY 142 CTATCTGGAGCGGGGATCCCTCCAGAACAGCCCGACAGGAGCCCTAGACAGCAGCT 201
Db 98 -----GlyGlyGlyGlyProGlnArgProGlyProSerProArg----- 112
QY 202 CCCACCATCCCAAGGCCCTGCGCCACGACACCATGTCGCCGAGACCAAGTGCCTTTAC 261

Db 113 ProLeuValProAlaGlyProAlaPro----- 122
QY 262 TCTGCTCTCTCCAAAGCGGTCTCTTCCCGGAGGACCAGAGAGGAGGAGGAGGAGTGTCT 321
Db 123 -----AlaAlaLysLeu-----ArgProProGluGlySerAlaGlyAla 136
QY 322 -----GAACCATGT----- 330
Db 137 CysAlaProValProAlaAlaAlaAlaAlaGlyProGluProAlaGly 156
QY 331 -----CCTAAGGACATTGAGCTGTTCATGGAAAGCTGGAGAGGCCAGGCAAA 381
Db 157 ProAlaLysProAlaGlyProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 176
QY 382 GACCAGCAGGAAGAAATTTGGGAAAAAACAACAGGAGGAGGAGTCTCACCAGGC 441
Db 177 Gly-----ProGlyProGlyProGly----- 185
QY 442 ACAGTACATTGACTGCTTCCAGAAGATCAAGTACAGCTTCAACCTCTGGGAGGCTGCG 501
Db 186 -----LysProAlaGlyProGlyAla 192
QY 502 CACCTGGCTGAAGGAGACAAAGTGCCTGAGCTGCTACATCCTCTTCAAGTCCCTGAA 561
Db 193 AlaGlnThrLeuAsnGlySerAlaAlaLeuLeuAsnSerHisHisAlaAlaProAla 212
QY 562 CTTTCATCTCTGGCAGGTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 618
Db 213 ValSerLeuValAsn-----AsnGlyProAlaAlaAlaLeuProLeu 226
QY 619 CCTCACCCCTAAAGCTATCAACCTGCTACAGTCTCTCTAAGCCCACTGAGAGTAACT 678
Db 227 ProLysPro-----AlaAlaProGlyThrValIleGlnThrPro-----Pro 240
QY 679 TTGGATGGGTGG-GCCAGGCTGGACCACTAGCC-----GGG 716
Db 241 PheValGlyAlaAlaAlaProProAlaProAlaAlaAlaProSerProProAlaAlaProAla 260
QY 717 CCGACTGGACAGCGGATGAGCCCTGCCCTACCACCCACAT-----TCTCAG 764
Db 261 ProAlaAlaProAlaAlaAlaAlaProProProProProProAlaProAlaThrLeuAlaArg 280
QY 765 ATGACTGGCAACTTCCAGAGCCCTCCAGCAAGCAGCCTTAGCATACAGGACCCCTGTTT 824
Db 281 ProGlyHisProAlaGlyProProThrAlaAlaPro-AlaValProProProAlaAl 300
QY 825 CCTTCGCGGGGAGTCAAGTCTTGTACGAGTTTCAAGCTAGGAACCCAGGGAAGTGA 884
Db 300 aAlaGlnAsnGlyGly-----SerAlaGlyAlaAlaPr 311
QY 885 ACAACCATGACCTCAGCTGAGCCCTGGGAGCCCACTCCAGGCGCTCCAGCCCAACCTGCC 944
Db 311 oAlaProAlaProAlaAlaAlaGlyProAlaGlyValSerGlyGlnProGlyProGlyAl 331
QY 945 AGCCAGCCCTGAAATGCAAGTCTTGTACGAGTTTCAAGCTAGGAACCCAGGGAAGTGA 1004
Db 331 aAlaAlaAlaAlaProAlaProGlyValLysAlaGluSerProLysArgValValGlnAl 351
QY 1005 CTGTGCTCCAGGAGAGAAAGCT-----GGAGGTTCTGGACCAACAGCGGTGGTGTCTG 1061
Db 351 aAlaProProAlaAlaGlnThrLeuAlaAlaSerGlyProAlaSerThrAlaAlaSer-M 371
QY 1062 TGAAGAAATGAGGCGGAGGAGGCGGTACATTCCAAAGC---AACATCTGGGAGCCCTTAC 1118
Db 371 etValIleGlyProThrMetGlnGlyAlaLeuProSerProAlaAlaValProProAla 391
QY 1119 AGCCGGGAGCCCT-----GGGACCCAGGCGGCTGACCT 1154
Db 391 laProGlyThrProThrGlyLeuProLysGlyAlaAlaGlyAlaValThrGlnSerLeu 411
QY 1155 CTCGGGTTCCAATG-----CTTCGACTTAGCTCGAGGCGCTGAAGAGGTCA 1199


```
Qy 1070 -----GAGCGGGAGCGGCTTACATTCCAAGCAACAT 1104
      :::::  :::::  :::::  :::::  :::::
Db 866 oTyAspSerAsnArgValValLeuArgSerGlyLysAspTyrIleAsnAlaSerCy 886
Qy 1105 CTTGAGCCCTTACAGCGGGGCCCT-----GGGACCGAGCGGCTACACCC 1153
      :::::  :::::  :::::  :::::  :::::
Db 886 sValgluglyLeuSerProTyrCysProLeuValAlaThrGlnAlaProLeuPro 905
      :::::  :::::  :::::  :::::  :::::
RESULT 9
US-09-735-367B-3
; Sequence 3, Application US/09735367B
; Patent NO. US20020151477A1
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Jan-Ake
; APPLICANT: Caltra, Françoise
; APPLICANT: Antonsson, Per
; TITLE OF INVENTION: NUCLEAR RECEPTOR COACTIVATOR
; FILE REFERENCE: 102093-100
; CURRENT APPLICATION NUMBER: US/09/735, 367B
; CURRENT FILING DATE: 2000-12-12
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 2005
; TYPE: PRT
; ORGANISM: mammal
US-09-735-367B-3

Alignment Scores:
Pred. No.: 0.000424 Length: 2005
Score: 164.00 Matches: 131
Percent Similarity: 33.51% Conservative: 60
Best Local Similarity: 22.98% Mismatches: 207
Query Match: 5.24% Indels: 173
DB: 10 Gaps: 27

US-09-762-021A-1 (1-1710) x US-09-735-367B-3 (1-2005)
Qy 32 GCTCTGAGGAAGCTGGAGCAAGACCTCGACTTGGAGCCCTTACGCGCCAGGAC 91
      :::::  :::::  :::::  :::::  :::::
Db 897 SerLeuAspAsnSerGlyProLysLeuProGluPheSerAsnArgProGlyTyrPro 916
Qy 92 AGATGGAGGGGCTGTATGGAAGCCGCTCCCTATGAGGAGGACGCTATCTGGAG 151
      :::::  :::::  :::::  :::::  :::::
Db 917 -----SerGlnProValGluInArgProLeuGlnMetProGlnLeuMetGln 934
Qy 152 CCGGGGATCCCTCAGAACACGCCACAGAGGACCCCTAGAGCACAGCCTCCACCATCC 211
      :::::  :::::  :::::  :::::  :::::
Db 935 HisValAlaProProGlnProGlnProGlnProGlnProGlnProGlnProGln 954
Qy 212 CCAAGGCCCTGCCACGCCAC-----ACCAGTCCCGAGAACCAAGTGCCCTTT 232
      :::::  :::::  :::::  :::::  :::::
Db 955 GlnGlnProProProProSerGlnProGlnSerGlnGlnGlnGlnGlnGlnGln 974
Qy 233 -----ACCAGTCCCGAGAACCAAGTGCCCTTT 259
      :::::  :::::  :::::  :::::  :::::
Db 975 MetMetMetMetMetMetGlnInAspProLysSerValArgLeuProValSerGln 994
Qy 260 ACTCTGCTCTCCAAAGCGGTCTCTTCCCGGAGGACCCAGAGAGGGACGAGGAAGTG 319
      :::::  :::::  :::::  :::::  :::::
Db 995 AsnValHisProProArg-GlyProLeuAsnPro----- 1005
Qy 320 CTGNACCATGTCCTAAGGACATTGACTGTTCTATGGGAAAGCTGGAGAGGCCAGGCA 379
      :::::  :::::  :::::  :::::  :::::
Db 1006 -----AspSerGlnArgMetProMetGln 1013
Qy 380 AAGACGAGGAGGAAGAAATTGGGAAAAAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 439
      :::::  :::::  :::::  :::::  :::::
Db 1013 nGlnSerGly-SerValProValMetValSerLeuGlnGlyProAlaSerValProPro 1033
Qy 440 GCACAGTACATTGACTGCTTCCAGAGATCAAGTACAGCTTCAACCTCCT----- 489
```

```
Db 1033 er-----ProAspLysGlnArgMetProMetProValAsnThrProL 1047
      :::::  :::::  :::::  :::::  :::::
Qy 490 --GGGAGGCTGGCCACCTGGCTCAAGAGAGACAAGTCCCTCGAGCTCGTACACATCCTC 547
      :::::  :::::  :::::  :::::  :::::
Db 1047 euGlySerAsnSerArgLysMetValTyrGlnGluSerPro-----GlnAsnProS 1064
Qy 548 TTCAAGTCCCTGAACTTCATCTCGCCGAGGTGCCCTGAGGCT---GGCTAGCAGCCCAA 604
      :::::  :::::  :::::  :::::  :::::
Db 1064 erSerSerProLeuAlaGlu-MetAlaSerLeuProGluAlaSerGlySerGluAlaPro 1083
Qy 605 GTGATCTCACCCCTCTCACCCCTAAAGCTATCAACCTGCTACAGTCTGTCTAAGCCCA 664
      :::::  :::::  :::::  :::::  :::::
Db 1084 SerVal-----ProGlyGlyProAsnAsnMetProSerHisValValLeu 1098
Qy 665 CTTGAGAGTAACCTTTTGGATGGGGTGGCCACCTGGACCACTACCGCGGCGGCTGG 724
      :::::  :::::  :::::  :::::  :::::
Db 1099 ProGlnAsnGlnLeuMetMet----- 1105
Qy 725 ACAGGCGATGAGCCCTGCTCCCTACCAACCCACATCTCTCAGATGACTGGCAATCTCCAGAG 784
      :::::  :::::  :::::  :::::  :::::
Db 1106 ThrGlyProLysProGlyPro-----SerProLeu 1115
Qy 785 CCTTCCAGCCCAAGCACCCCTTAGGATACCAAGACCCCTGTTCCCTCGCGGGGAGTCA 844
      :::::  :::::  :::::  :::::  :::::
Db 1116 SerAlaThrGlnGlyAlaThrProGlnGlnProProValAsnSerLeuProSerHis 1135
Qy 845 AGTTGAGGAGCACCTCACACTTTCCT-----CAGGAG 877
      :::::  :::::  :::::  :::::  :::::
Db 1136 -----GlyHisHisPheProAsnValAlaAlaProThrGlnThrSerArgPro 1151
Qy 878 AAGACACACAACCAT---GACCCCTCAG-----CCTGGGACCCCACTCCAGGCC 925
      :::::  :::::  :::::  :::::  :::::
Db 1152 LysThrProAsnArgAlaSerProArgProTyrProGlnThrProAsnAsnArgPro 1171
Qy 926 TCCAGCCCCAAACCTGCCAGCCAGCCCTGAAAAATGCAAGTCTTGTACGAGTTTGAAGCT 985
      :::::  :::::  :::::  :::::  :::::
Db 1172 ProSerThrGluProSerGluIleSerLeu----- 1181
Qy 986 AGNACCCAGCGGNACTGACTGTGTCAGGAGAGAGCTGGAGGTTCTGGACCACAGC 1045
      :::::  :::::  :::::  :::::  :::::
Db 1182 ---SerProGluArgLeu----- 1186
Qy 1046 AAGCGGTGGTGGTGAAGAATGAGCGGACGAGCGGTACATTTCCAAGCAACATC 1105
      :::::  :::::  :::::  :::::  :::::
Db 1187 -----AsnAlaSerIleAlaGlyLeuPheProGlnIle 1198
Qy 1106 CTGAGGCCCTTACAGCGGGACCCCTGGG-----ACCAGGGCCAG 1147
      :::::  :::::  :::::  :::::  :::::
Db 1199 AsnIleProLeuProPro---ArgProAsnLeuAsnArgGlyPheAspGlnGlnGlyLeu 1217
Qy 1148 TCACCTCTCGGGTTCGAATGCTTCGACTTAGCTCGAGGCTGAGAGGCTCAGAGCTGG 1207
      :::::  :::::  :::::  :::::  :::::
Db 1218 AsnProThr-----ThrLeuLysAlaIleGlyGlnAlaProSerAsnLeuThr- 1233
Qy 1208 CTGAGGAGAGAACTTCTCCACTGCCCGTGGAGGACACTTGGTCCCTGAC----- 1260
      :::::  :::::  :::::  :::::  :::::
Db 1234 MetAsnProSerAsnPheAlaThrProGlnThrHisLysLeuAspSerValValAsn 1253
Qy 1261 ---GGGAGCGCCACTTCTCGCTAAGACCTGGGAGCTACAGATGT----- 1305
      :::::  :::::  :::::  :::::  :::::
Db 1254 SerGlyLysGlnSerAsnSerGlyAlaThrLysArgAlaSerProSerAsnSerArg 1273
Qy 1306 ATGTCCACAGGAGGCCCCCAAGATCTCTCCGCTGGAGGCTGTCAAGAGGATGCTGGG 1365
      :::::  :::::  :::::  :::::  :::::
Db 1274 SerSerProGlySerSerArgLysThrProSerProGlyArgGlnAsnSerLysAla 1293
Qy 1366 GATAAGCCCTTAGGCACAGCTTAGACACCTCCAAGAACAGGCCGCTCATCAAGAT 1425
      :::::  :::::  :::::  :::::  :::::
Db 1294 ProLys-----LeuThrLeu---AlaSerGlnThrAsnAlaAlaLeuLeuGlnAsn 1309
Qy 1426 GGCAGATCTGATACCCCATTAGAGCCCGAGAAATTCCTTCTTGTGATCCCACT---TTGCA 1482
      :::::  :::::  :::::  :::::  :::::
```

Db 1310 -ValGluLeu-----ProArgAsnValLeuValSerProThrProLeuAl 1324
Qy 1483 GCAAAACCCACACCCAGCTCACACAGCAAAACAAATGGACAGGCCAGAGGCTGAAGCA 1542
Db 1324 aAsnProValProGlySerPheProAsnAsnSerGlyLeuAsnProGln---AsnSe 1343
Qy 1543 AACAGTGTCCTTCTGGCTGTG 1564
Db 1343 rThrValSerValAlaAlaVal 1350
RESULT 10
US-09-735-367B-2
; Sequence 2, Application US/09735367B
; Patent No. US2002015147A1
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Jan-Ake
; APPLICANT: Calra, Françoise
; APPLICANT: Anonsson, Per
; TITLE OF INVENTION: NUCLEAR RECEPTOR COACTIVATOR
; FILE REFERENCE: 102093-100
; CURRENT APPLICATION NUMBER: US/09/735,367B
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/174,544
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2063
; TYPE: PRT
; ORGANISM: Human
US-09-735-367B-2

Alignment Scores:

Pred. No.:	0.000504	Length:	2063
Score:	163.00	Matches:	131
Percent Similarity:	33.51%	Conservative:	60
Best Local Similarity:	22.98%	Mismatches:	207
Query Match:	5.21%	Indels:	173
DB:	10	Gaps:	27

US-09-762-021a-1 (1-1710) x US-09-735-367B-2 (1-2063)

Qy 32 GCTGTGGAGGAGAGCTGGAGCAAGACTCTGACTTGGAGGCTTCAGCCAGGCCAGGAC 91
Db 955 SerLeuAspAsnSerGlyProLysLeuProGluPheSerAsnArgProGlyTyrPro 974
Qy 92 AGATGGAGGGGGCTGCTATGGAAGGCCGCTCCCTATGGAGCAGGCAGCTATCTGGAG 151
Db 975 -----SerGlnProValGluGlnArgProLeuGlnGlnMetProProGlnLeuMetGln 992
Qy 152 CCGGGGATCCCTCCAGAACAGCCCCCAGAGGACCTTAGAGCAGAGCTCCACCATCC 211
Db 993 HsValAlaProProGlnProGlnProGlnGlnGlnProGlnProGlnLeuProGlnGln 1012
Qy 212 CCAAGGCCCTGCCAGGCCAC----- 232
Db 1013 GlnGlnProProProSerGlnProGlnSerGlnGlnGlnGlnGlnGlnGlnGln 1032
Qy 233 -----ACCACTGCCGAGAGAACCAAGTGCCTTT 259
Db 1033 MetMetMetMetLeuMetMetGlnAspProLysSerValArgLeuProValSerGln 1052
Qy 260 ACTGTCCTCTCCAAGCGGTCTCTTCCCGGAGGCCAGAGGCCAGAGGAGGAGGAGTG 319
Db 1053 AsnValHisProProArg-GlyProLeuAsnPro----- 1063
Qy 320 CTGAACCATGCTCAAGGACATTGAGCTGTTCATGGGAAAGCTGGAGAGGCCAGGCA 379
Db 1064 -----AspSerGlnArgMetProMetG1 1071
Qy 380 AAGACCACGAGAGCAAGAAATTTGGGAAAAAACAAGGAGGCCAGGAGTCTCACCCAG 439
Db 1071 nGlnSerGly-SerValProValMetValSerLeuGlnGlyProAlaSerValProPros 1091


```
QY 1426 GGCAGATCTGATACCCATTAGAGCCCGAGAAATTCCTCTTGTGATCCCACT---TTGCA 1482
    ::::: ||||| ||::: ||::: ||
Db 1368 -ValGluLeu-----ProArgAsnValLeuValSerProThrProLeuAl 1382
    ::::: ||||| ||::: ||::: ||
QY 1483 GCRAACCCACACCCAGCTCACACACCAAAACAATGGACAGCCAGAGGCTGAAGCA 1542
    ||| ||| ||| ||::: |||
Db 1382, aAsnProProValProGlySerPheProAsnAsnSerGlyLeuAsnProGln---AsnSe 1401
    ||||| ||::: |||
QY 1543 AACAGTCTCCCTTCTGGCTGTG 1564
    ||||| ||::: |||
Db 1401 rThrValSerValAlaVal 1408
    ||||| ||::: |||

RESULT 11
US-09-858-664A-3
; Sequence 3, Application US/09858664A
; Patent No. US20020072491A1
; GENERAL INFORMATION:
; APPLICANT: WEI, Ming-Hui, et al.
; TITLE OF INVENTION: ISOLATED HUMAN KINASE PROTEINS, NUCLEIC
; MOLECULES, ACID MOLECULES ENCODING HUMAN KINASE PROTEINS, AND USES
; FILE REFERENCE: CL000927-CIP
; CURRENT APPLICATION NUMBER: US/09/858, 664A
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/711,134
; PRIOR FILING DATE: 2000-11-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 846
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-858-664A-3

Alignment Scores:
Pred. No.: 0.000526 Length: 846
Score: 161.50 Matches: 128
Percent Similarity: 29.56% Conservative: 39
Best Local Similarity: 22.65% Mismatches: 167
Query Match: 5.22% Indels: 231
DB: 10 Gaps: 27

US-09-762-021A-1 (1-1710) x US-09-858-664A-3 (1-846)
QY 1490 GGGTTTCTGCAAACTGGATCCAGAGAGAAATTCCTCGGGCTCTAATGGGTATCAGAT 1431
    ||| ::::: ||| ||||| ||::: ||::: ||
Db 433 GlyLeuSerSerSerSerAspSerGluGluGlu-----GluLeuGluGlu 447
    ||||| ||::: |||
QY 1430 CTGCCATCTTGCATPCAGCGGGGCTGTCTTCTGGAGGTGTCTAAGCTGGTGCCTAAGGC 1371
    ||||| ||::: |||
Db 448 LeuProSer-----ValProArgPro 454
    ||||| ||::: |||
QY 1370 TTATCCCCAGCATCTTCTGACAGCTCCAGCCGGGACAGATTCTGGGGCTCCTGTG 1311
    ||| ||| ||| ||| |||
Db 455 LeuGlnProGlu-----PheSerGlySerArgVal 464
    ||||| ||::: |||
QY 1310 GACATAGCATCTGTAGTCCCGAGTCTTATGCAAGTAGTGGCTCCCG---TCAGGG 1254
    ::::: ||| ||| ||| ||| |||
Db 465 SerLeuThrAspIleProThrGlu-----AspGluAlaLeuGlyThrProGluThrGly 482
    ::::: ||| ||| ||| ||| |||
QY 1253 ACCCAAGTCTCTACCGTGGCAGTGGAGAGATTTCTCTGCTGCAGCCAGTCTGTGACCT 1194
    ||||| ||::: |||
Db 483 AlaAlaThrProMetAspTrpGlnGluGlnGlyArgAlaPro----- 496
    ||||| ||::: |||
QY 1193 CTTCAGGCTCGAGCTAAGTCGAGCATTTGGAACCCAGAGGTGACTGGCCCTGGGTCC 1134
    ||||| ||||| ||||| |||
Db 497 -----SerGlnAspGlnGluAlaProSerProGluAlaLeuProSerProGlyGln 513
    ||||| ||||| ||||| |||
QY 1133 CAGGGTCCCGGCTGTAGGGGCTCCAGGATGTGCTTGGAAATGTAGCCGC----- 1083
    ::::: ||||| ||||| |||
Db 514 Glu-----ProAlaAlaGlyAlaSerProArgArgGlyGluLeuArgArgGlySerSer 531
    ::::: ||||| ||||| |||
QY 1082 -----TCCGTCCCGCTCATCTTCTTCCAGCCACC-----ACCGCT 1047
    ::::: ||||| ||||| |||
```

```
Db 532 AlaGluSerAlaLeuProArgAlaGlyProArgGluLeuGlyArgGlyLeuHisLysAla 551
    ||| ||| ||| ::::: |||
QY 1046 TGTGTGTGGTCCAGAACCTCCAGCTTCTCTCCCTGACACAGTCCAGTTCCTCCGTGGTTC 987
    ||| ||| ||||| ||::: |||
Db 552 AlaSerValGluLeuProGlnArgArgSerProGlyProGlyAla----- 566
    ||||| ||||| ||||| |||
QY 986 TAGCTTCAAACTCTACAAAGACTTTCATTTTTCAGGGTGGCTGGCCAGGTTTGGGGCTGG 927
    ||||| ||||| ||||| |||
Db 567 -----ThrArgLeuAlaArgGlyLeu----- 574
    ||||| ||||| ||||| |||
QY 926 AGGCCCTGGAGTTGGGGTCCCGAGGCTGAGGGTGTGCTGTCTTCTCTCCTGAGGAA 867
    ||| ||| ||| |||
Db 575 -----GlyGluGlyGluTyraAlaGlnArgLeuGln 584
    ||||| ||||| ||||| |||
QY 866 AGTGTAGGTGCTCCCTAACCTATGACTTCCCGCCCGGAGGAAA-----CAG 819
    ::::: ||||| ||| ||| ||||| |||
Db 585 AlaLeuArgGlnArgLeuArgLeuArgGlyGlyProGluAspGlyLysValSerGlyLeuArg 604
    ||||| ||||| ||||| |||
QY 818 GGTCTGTATCTCCCTAAGGGTCTTGGCTGGAGGCT----- 783
    ||||| ||||| ||||| |||
Db 605 GlyPro---LeuLeuGluSerLeuGlyGlyArgAlaArgAspProArgMetAlaArgAla 623
    ||||| ||||| ||||| |||
QY 782 -----CTGGAAGTTGCCAGTCA----- 765
    ||| ||| ||| |||
Db 624 AlaSerSerGluAlaAlaProHisHisGlnProProLeuGluAsnArgGlyLeuGlnLys 643
    ||||| ||||| ||||| |||
QY 765 ----- 765
    ||||| ||||| ||||| |||
Db 644 SerSerSerPheSerGlnGlyGluAlaGluProArgGlyArgHisArgArgAlaGlyAla 663
    ||||| ||||| ||||| |||
QY 764 ---CTGAGAATGTGGTGTAGGGAGGGCTCATCGCTGTCCAGTCCAGTCCCGCGGTAG 708
    ||| ||| ||| ||||| ||||| |||
Db 664 ProLeuGluIleProValAlaArgLeuGlyAlaArgArgLeuGlnGluSerPro----- 681
    ||||| ||||| ||||| |||
QY 707 TGTCTCAGGCTGGGCCCAACCCCATCAAGGTTACTCTCAGGTGGGCTTACAGAGGACT 648
    ||||| ||||| ||||| |||
Db 681 ----- 681
    ||||| ||||| ||||| |||
QY 647 GTACGAGTTGATAGCTTAGGGGTGAGAGGGGTGAGATCAGTCTGGCTCTAGGCAG 588
    ||||| ||||| ||||| |||
Db 682 -----SerLeuSerAlaLeuSerGlu 688
    ||||| ||||| ||||| |||
QY 587 CCTCAGGCACCTGCCAGGATGAAGTTACAGGACTTGAAGAGGATGTGTACGAGCTCAG 528
    ||| ||| ||| ||| |||
Db 689 AlaGlnProSerSerProAla----- 695
    ||||| ||||| ||||| |||
QY 527 GGGCAGTTGTCTCTCAGCCAGGCTGGCCAGCCTTCCAGAGGAGTTGAAGCTGTACTTGA 468
    ||||| ||||| ||||| |||
Db 696 -----ArgProSerAlaProLysProSerThrPro----- 705
    ||||| ||||| ||||| |||
QY 467 TCTTCTGGAAGCAGTCAATGTACTGTGCTGGGTGAGACCTCCCTGGTCTGTTGTTTTTT 408
    ||| ||| ||| ||| ||| |||
Db 706 LysSerAlaGluProSerAlaThrThrProSer---AspAlaProGlnPro----- 721
    ||||| ||||| ||||| |||
QY 407 TCCCAAAATTTCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 348
    ||| ||| ||| ||| ||| |||
Db 722 -----ProAlaProGlnProAlaGlnAsp---Lys 730
    ||||| ||||| ||||| |||
QY 347 GCTCAATGTCTCTTAGGACATGTTACAGCAGTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 288
    ||| ||| ||| ||| ||| |||
Db 731 AlaProGluProArgProGluProValArgAlaSerLysProAlaProProGlnAla 750
    ||||| ||||| ||||| |||
QY 287 AAGAGGACCCCTTGGAGGAGGAGAGTAAAGGACACTTGGTCTCGGGCACTGGTGTGGC 228
    ::::: ||||| ||||| |||
Db 751 LeuGlnThrLeuAla-----LeuProLeuThrProThrAlaGln 763
    ||||| ||||| ||||| |||
QY 227 GTGCAGGGGCTTGGGAGGCTGGTGGAGGCTGTGCTCTAGGGTCTCTTGGTGGGCTGTT 168
    ::::: ||||| ||||| |||
Db 764 IleIleGlnSerLeu---GlnLeuSerGlyHisAlaGlnGlyProSer----- 778
    ::::: ||||| ||||| |||
QY 167 CTGAGGAGTCCCGGCTCCAGATAGCGTCTCTCCATAGGAGGCGGCTTT----- 114
    ::::: ||||| ||||| |||
```

```
Db 779 ----GInGlyProAlaAlaProSerGluProLysProHisAlaAlaValPheAlaArg 797
Qy 113 -----CCATAGCAGGCCCCCTCCATC-----TGTCCT 87
Db 798 ValAlaSerProProGlyAlaProGluLysArgValProSerAlaGlyGlyProPro 817
Qy 86 GGCCTGGCTGAAGGCTCCAAAGTCAGAGTCTTTGCTCCAGCTCTTCTCCAGAGCTTCT 27
Db 818 ValLeuAlaGluLysAlaArgVal-----ProThrValProProArgProGly 833
Qy 26 GCAGCTGGTCTTCA 12
Db 834 SerSerLeuSerSer 838
RESULT 12
US-09-788-043C-5
; Sequence 5, Application US/09788043C
; Patent No. US20020107361A1
; GENERAL INFORMATION:
; APPLICANT: Zuo, Fengrong
; APPLICANT: Klonowski, Paul
; TITLE OF INVENTION: No. US20020107361A1el Metalloproteases Having
; TITLE OF INVENTION: Thrombospondin Domains and Nucleic Acid Compositions
; TITLE OF INVENTION: Encoding the Same
; FILE REFERENCE: ROCH-004
; CURRENT APPLICATION NUMBER: US/09/788,043C
; CURRENT FILING DATE: 2001-02-16
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1690
; TYPE: PRT
; ORGANISM: human
US-09-788-043C-5
Alignment Scores:
Pred. No.: 0.00167 Length: 1690
Score: 155.50 Matches: 133
Percent Similarity: 30.32% Conservative: 38
Best Local Similarity: 23.58% Mismatches: 176
Query Match: 4.97% Indels: 217
DB: 10 Gaps: 32
US-09-762-021a-1 (1-1710) x US-09-788-043C-5 (1-1690)
Qy 142 CTATCTGGAGCCGGGATCCCTCCAGAACAGCCCCAGAGGAGCCCTAGAGCAGGCT 201
Db 1092 LeuAlaGlyThrGlyAsp-----ArgThrProProHisSerArgProAlaAlaPro 1109
Qy 202 CCCACATCCCAAGGCCCTCCAGGCCACACAGTCCCGAGAACCAAGTGCCTTTAC 261
Db 1110 SerThrGlySerProValProAlaThr-----GluProPro----- 1121
Qy 262 TCTGCTCTCTCAAGCGCTCTCTTCCCGGAGGACCCAGAGGAGGAGGAGTGCCT 321
Db 1122 AlaAlaLysGluGluGlyValLeuGlyProTrpSerPro----- 1134
Qy 322 GAACCATGTCTTAAGGACATTGAGCTGTTCATGGGAAGCTGGGAAGGCCAGGCAAA 381
Db 1135 SerProTrpPro-----SerGlnAlaGlyArgSerProProPro 1147
Qy 382 GACCAGCAG-----GAAGACAAATT----- 402
Db 1148 ProSerGluGlnThrProGlyAsnProLeuIleAsnPheLeuProGluGluAspThrPro 1167
Qy 403 -----TGGGAAAAAACAAGGACCCAGG 426
Db 1168 IleGlyAlaProAspLeuGlyLeuProSerLeuSerTrpProArgValSerThrAspGly 1187
Qy 427 AGGTCTCACCCAGGCACAGTACATTGACTGTCTCCAGAA----- 465
```

```
Db 1188 LeuGlnThrProAlaThr-----ProGluSerGlnAsnAspPheProVal 1202
Qy 466 -----GATCAAGTACAGCTTCAACCTCCT-----GGGAAGGCTGGCCACCTGGCT 510
Db 1203 GlyLysAspSerGlnSerGlnLeuProProTrpArgAspArgThrAsnGluValPhe 1222
Qy 511 GAAGGAGACAAGTGCCTGAGCTCGTACACATCCTCTTCAAGTCCCTGAACATTCTATCCT 570
Db 1223 LysAspAspGluGluProLysGlyArgGlyAlaPro----- 1234
Qy 571 GGCCAGGTGCCCTGAGCTGAGCTAGCAGCCCAAGTGTATCTACCCCTCTCCACCCCTAA 630
Db 1235 ---HisLeuPro-----ProArgProSerSerThrLeuProPro-LeuSerPro-- 1249
Qy 631 AGCTATCAACCTGCTACAGTCTCTTAAGCCCACTCAGAGTACCTTTTGATGGGCTT 690
Db 1250 ---ValGlySerThrHisSerSerProSerProAspValAlaGluLeuTrpThrGlyGl 1268
Qy 691 GGGCCAGCCTGGACCACTAGCCGGCGGCTGGACAGCGCATGAGCCCTGCCCTAGCA 750
Db 1268 yThrValAlaTrpGluProAlaLeuGluGlyGlyLeuGlyProValAspSerGluLeuTr 1288
Qy 751 ACCCACATCTCTCAGATGACTGGCAACTTCCAGAGCCCTCCAGCCCAAGCACCTTTAGGATA 810
Db 1288 pProThrValGlyValAlaSerLeuLeuProProPhe-----AlaProLeu----- 1304
Qy 811 CCAGGACCTGTTCCTTCCTTCGGCGGGAAGTACATAGTTAGGGAGCACCTCACACTTCC 870
Db 1305 ----- 1305
Qy 871 TCAGGAGAGACACACAACCACTGACCCCTCAGCTGGGACCCCAACTCCAGGCCCTCCAG 930
Db 1305 oGluMetLysValArgAspSerSerLeuGluProGly-----ThrProSerPh 1321
Qy 931 CCCAAACCTGCCAGCCAGCCCTGAAATGCAAGTCTTTGTACGAGTTTGAAGCTAGGAA 990
Db 1321 eProAlaProGlyProGlySerTrpAspLeuGlnThrValAlaValTrpGlyThrPheLe 1341
Qy 991 CCCCGGGAAGTACTGTGCTCCAGGGAGAGAGCTGGAGGTTCTGGACCACAGCAAGCG 1050
Db 1341 uProThrThrLeuThrGlyLeu----- 1348
Qy 1051 GTGGTGGCTGGTGAAGAAATGAGCGGGGACGAGCGGCTACATTTCCAAGCAACATCCTGA 1110
Db 1349 -----GlyHisMetProGluProAlaLeu-- 1356
Qy 1111 GCCCTACAGCCGGGACCCCTGGGACCCAGGCGGACGCTCCTCGGTTTCCAATGCT 1170
Db 1357 -----AsnProGlyProLysGlyGlnProGluSerLeuSerProGluVa 1371
Qy 1171 TCGACTTAGCTCGAGGCTGGAAGAGGTCACAGACTGG-----CT 1209
Db 1371 iProLeuSerSerArgLeuLeuSerThrProAlaTrpAspSerProAlaAsnSerHisAr 1391
Qy 1210 GCAGGCACAGAACTTCTCCACTG---CCACGGTGGAGGACACTTGGGCTCCC----- 1256
Db 1391 gValProGlu-ThrGlnProLeuAlaProSerLeuAlaGluAlaGlyProProAlaAspP 1411
Qy 1257 -----TGACGGGGGAGCCAGCTACTTCCGATAGAC----- 1286
Db 1411 roLeuValValArgAsnAlaSerTrpGlnAlaGlyAsnTrpSerGluCysSerThrThrc 1431
Qy 1287 -----CTGGGAGCTAC----- 1298
Db 1431 ysGlyLeuGlyAlaValTrpArgProValArgCysSerSerGlyArgAspGluAspCysA 1451
Qy 1299 -----AGATGCTATGTCACAGAGGCGCCACGAAT 1329
Db 1451 laProAlaGlyArgProGlnProAlaArgArgCys-----HisLeuArgPro----- 1466
Qy 1330 CCTGTCCCGGCTGG-----AGCTGTCAAGAGGATGCTGGGGATANG 1371
```

Db 1467 --CysAlaThrTrpHisSerGlyAsnTrpSerLysCysSerArgSerCysGlyGly---- 1484
QY 1372 CCCTTAGCCACCACTTAGACACCTCCAGAACACGAGCCCGCTGATGCAAGATGGCAGA 1431
Db 1485 -----GlySerSerValArgAspValGlnCysValAspThrArgAspLeuArg----- 1500
QY 1432 TCTGATACCCATTAGACCCCGAGAAATTCCTCTCTCGATCCCAAGTTTGGACCAACCC 1491
Db 1501 -----ProLeuArgPro-----PheHisCysGlnProGlyProAlaLysPro- 1514
QY 1492 ACACCCCGCTCACACAGCAAAACAATGACAGGCCCGCAGAGGCTGAAGCAAAACAGTGTC 1551
Db 1515 -----ProAlaHis-----ArgProCysGly---AlaGlnProCysL 1526
QY 1552 CCTTCTGG 1559
Db 1526 euSerTrp 1528

RESULT 13
US-09-919-497-56
; Sequence 56, Application US/09919497
; Patent No. US20020106662A1
; GENERAL INFORMATION:
; APPLICANT: Mutter, George L.
; TITLE OF INVENTION: PROGNOSTIC CLASSIFICATION OF ENDOMETRIAL CANCER
; FILE REFERENCE: B0801/77225
; CURRENT APPLICATION NUMBER: US/09/919,497
; CURRENT FILING DATE: 2001-07-31
; PRIOR APPLICATION NUMBER: US 60/221,735
; PRIOR FILING DATE: 2000-07-31
; NUMBER OF SEQ ID NOS: 100
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 1806
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (758)..(758)
; OTHER INFORMATION: Xaa = any amino acid
; NAME/KEY: UNSURE
; LOCATION: (809)..(809)
; OTHER INFORMATION: Xaa = any amino acid
US-09-919-497-56

Alignment Scores:
Pred. No.: 0.002 Length: 1806
Score: 154.50 Matches: 166
Percent Similarity: 31.52% Conservative: 31
Best Local Similarity: 26.56% Mismatches: 229
Query Match: 4.94% Indels: 200
DB: 10 Gaps: 40

US-09-762-021A-1 (1-1710) x US-09-919-497-56 (1-1806)
QY 1 GCGAGACCGACTGAAGACCAAGCTGCA-----GAAGGCTCTGGAGGAGAGCTGGAGCA 54
Db 1015 GlyileSerGlyLysAspGlyProAlaGlyLeuArgGlyPheProGlyGluArgGlyLeu 1034
QY 55 AAGACCTCGACTTGGAGGCTTCAGCCAGCCAGGACAGATGGAGGGGCTGCTATGGA 114
Db 1035 ProGlyAlaGlnGlyAlaProGlyLeuLysGlyGly-----GluGlyPro---GlnGly 1051
QY 115 AAGGCCCTCCTATGGAGCA---GGCAGCGCTATCTGGAGCCGGGATCCCTCCAGAAC 171
Db 1052 ProProGlyProValGlySerProGlyGluArgGlySerAlaGly-----Thr 1067
QY 172 GCGCCACAGAGGACCTAGACACAGCCT---CCCACCATCCCAAGGCCCTGCCACG 228
Db 1068 AlaGlyProIleGlyLeuArgGlyArgProGlyProGlnGlyProProGlyProAlaGly 1087
QY 229 CCACACCATGCGCCGAGAACCAAGTGCTTTACTCTGCCTCTCCCAAG----- 276

Db 1088 GluLysGlyAlaProGlyGluLysGlyProGlnGlyProAlaGlyArgAspGlyValGln 1107
QY 277 -----GCGGTCTCTTCCCGGAGGACCCAGAGAGGAGGACGAGGAAGT----- 318
Db 1108 GlyProValGlyLeuProGlyProAlaGlyProAlaGlySerProGlyGluAspGlyAsp 1127
QY 319 -----GCTGAACCATGTCTTAAGGACATTGAGCTGTTCATGGGAAGACGTGGA 366
Db 1128 LysGlyGluIleGlyGluProGlyGlnLysGly-----SerLysGlyGly 1142
QY 367 GAAGGC----- 372
Db 1143 LysGlyGluAsnGlyProProGlyProProGlyLeuGlnGlyProValGlyAlaProGly 1162
QY 373 -----CCAGGCAAGACCCAGCAGGAAGAAATTTGGGAAAAA 411
Db 1163 IleAlaGlyGlyAspGlyGluProGlyProArgGlyGlnGlnGlyMetPheGlyGlnLys 1182
QY 412 AACAAAGGACGAGGAGGTCTCACCCAGGACACAGTACATTGCTCTCCAGAAGATCAA 471
Db 1183 GlyAspGluGlyAlaArgGlyPheProGly-----ProProGlyPro 1196
QY 472 GTACAGCTTCAA-----CCTCTGGGAAGCTGCCACCTGGCTGAAGGAGAC 519
Db 1197 IleGlyLeuGlnGlyLeuProGlyProProGlyGlnLysGlyGlu-----AsnGlyAsp 1214
QY 520 AAGTGGCCCTGAGCTCGTACACATCTCTTCAAGTCCCTGAACCTTCCTGGCCAGGTG 579
Db 1215 ValGlyProItrpGlyProProGlyPro-----ProGly----- 1225
QY 580 CCCTGAGGCTGGCTAGCAGCCCAAGTGATCTACCCCTCTCACCCCTAAAGCTATCAA 639
Db 1226 ProArgGly---ProGlnGlyProAsnGlyAlaAspGlyProGlnGly-----Pro 1241
QY 640 CTGCTACAGTCTGTCTTAAGCCACCTGAGAGTAACTTTTGGATGGGTGGGCC----- 695
Db 1242 ProGlySerValGlySer-----ValGlyGlyValGlyGluLys 1254
QY 696 -----CAGCCTGGACCA----- 707
Db 1255 GlyGluProGlyGluAlaGlyAsnProGlyProProGlyGluAlaGlyValGlyGlyPro 1274
QY 708 -----CTAGCCGGCGGCTGACAGAGGAGGATGAGCCCTGCG 743
Db 1275 LysGlyGluArgGlyGluLysGlyGluAlaGlyProProGlyAlaAlaGlyProGly 1294
QY 744 CCTACCAACCAACATCTCTAGATGACTGGCAACTTCCAGAGCCCTCCAGCCCAAGCACCT 803
Db 1295 AlaLysGlyProProGlyAspGlyProLysGlyAsnProGlyProValGlyPhePro 1314
QY 804 TAGGATACCAAGGACCTGTTTCCCTTCG-----GCGGGGAAGTCATAGGTAG 851
Db 1315 GlyAsp-ProGlyPro-----ProGlyGluLeuGlyProAlaGlyGlnAspGlyVal-G 1332
QY 852 GGAGCCTCTCACACTTCTCAGGAGAGACACACACCATGACCT---CAGCTGGGG 908
Db 1332 lyGly-----AspLysGlyGluAspGlyAspProGlyGlnProGlyP 1346
QY 909 ACCCCNACTCCA-----GGCCCTCCAGCCCAACCTGCCAGCCAGCCCTGA 956
Db 1346 roProGly-ProSerGlyGluAlaGlyProProGlyPro-----ProGly----- 1360
QY 957 AAATGCAAGTCTTGTAGAGTTTGAAGCTAGGAACCCAGGGAAGTACTGTGTGCCAGG 1016
Db 1361 -----LysArgGly-ProProGlyAlaAlaGlyAlaGluG 1372
QY 1017 GAGA-----GAAGCTGGAGTTCTGGACCACACAGAACGGGT 1052
Db 1372 yArgGlnGlyGluLysGlyAlaLysGlyGluAlaGlyAlaGluGlyProProGlyLysTh 1392
QY 1053 G-----GTGGCTGGTGAAGATGAGCGGGGAGCGGCGGTACA 1091
Db 1392 rglyProValGlyProGlnGlyProAlaGlyLysPro-----GlyProGluGlyLeuAr 1410

Qy 1092 TTCAAGCAACATCCT-----GGAGCCCTTACAGCC-----GGGGA 1127
Db 1410 gGlyleProGlyProValGlyGluInGlyLeuProGlyAlaAlaGlyGlnAspGlyPr 1430
Qy 1128 CCCCTGGAGCCAGCCGAGTACCCCTCTCGGGTTCATGCTTCGAGTTCGAGGC 1187
Db 1430 oProGlyProMetGlyProGlyLeuProGlyLeuLysGlyAspProGlySerLysG1 1450
Qy 1188 CTGAGAGGTACACAGCTGGCTGCGAGGAGAGAGACTTCTCCACTCCACGGTGAGGAC 1247
Db 1450 yGluLysGlyHisProGlyLeuileGly-----LeuileGlyProProGlyGluInG1 1468
Qy 1248 TTGGCTCCTCTGAC---GGGAGCCAGCTACTTCCGATAGAGCTGGGAGCTACAGATGC 1304
Db 1468 yGluLysGlyAspArgGlyLeuProGlyThrGln-----GlySerProGlyAl 1484
Qy 1305 TATGTCCACAGAA-----GGCCCCAGCAATCCTCTC-----CGGCTGG 1343
Db 1484 aLysGlyAspGlyGlyleProGlyProAlaGlyProLeuGlyProProGlyProProG1 1504
Qy 1344 A-----GGCTGTACAGAGGATGCTGGGGATAAG-----CCCTTAGGCACACAG 1385
Db 1504 yLeuProGlyProGlnGlyProLysGlyAsnLysGlySerThrGlyProAlaGlyGlnLy 1524
Qy 1386 CTAGACACCTCCAAAGAACCCAGCCCGCTGATGCAAGATGCGAGATCTGATACCCATTA 1445
Db 1524 sGlyAspSerGlyLeu--ProGlyProProGlyPro-----ProGly 1537
Qy 1446 GAGCCCCAGAAATTCCTCTTCTGGATCCAGTTCGACAAACCCACACCCAGCTCAC 1505
Db 1538 ProProGlyGluValileGlnProLeuProLysSerLysLysThrArgArgHis 1557
Qy 1506 ACA 1508
Db 1558 Thr 1558
RESULT 14
US-09-735-367B-6
; Sequence 6, Application US/09735367B
; Patent No. US20020151477A1
; GENERAL INFORMATION:
; APPLICANT: Gustafsson, Jan-Ake
; APPLICANT: Calra, Francoise
; APPLICANT: Antonsson, Per
; TITLE OF INVENTION: NUCLEAR RECEPTOR COACTIVATOR
; FILE REFERENCE: 102093-100
; CURRENT APPLICATION NUMBER: US/09/735,367B
; CURRENT FILING DATE: 2000-12-12
; PRIOR APPLICATION NUMBER: US 60/174,544
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 1070
; TYPE: PRT
; ORGANISM: mammal
US-09-735-367B-6
Alignment Scores:
Pred. No.: 0.00193 Length: 1070
Score: 154.00 Matches: 142
Percent Similarity: 31.98% Conservative: 63
Best Local Similarity: 22.15% Mismatches: 239
Query Match: 4.92% Indels: 198
DB: 10 Gaps: 32
US-09-762-021a-1 (1-1710) x US-09-735-367B-6 (1-1070)
Qy 20 ACCCTGCAGAGGCTCTGGAGCAAGAGCTGGGCAAGACCTCTGAGCTGGAGCCCTTCAG 79
Db 134 AlaIleAsnLeuAlaLeuAlaGlnAsnArgSerGlnAspValArgMetAsnGlyProMet 153

Qy 80 -----CCAGGCCAG 88
Db 154 GlyAlaGlyAsnSerValArgMetGluAlaGlyPheProMetAlaSerGlyProGlyIle 173
Qy 89 CACAGATGGAGGGGGCTGCT-----ATGGAAGCCGCTCCCTATATGAGCAGCAGCAGC 142
Db 174 IleArgMetAsnAsnProAlaThrValMetIleProProGlyGlyAsnValSerSer 193
Qy 143 TATCTGGAGCCGGGATCCCTCCAGACACAGCCCCACAGAGGACCTTAGACACAGCCTC 202
Db 194 MetMetAlaProGlyProAsnProGluLeuGln----- 204
Qy 203 CCACATCCCCAAGGCCCTCCACGC----- 229
Db 205 ProArgThrProArgProAlaSerGlnSerAspAlaMetAspProLeuLeuSerGlyLeu 224
Qy 230 CACACCTAGTCCCGAGAACCAAGTGCCTTACTGTGCCTCTCCAGCGGTCTCTTCC 289
Db 225 HistleGlnGlnSerHisProSerGlySerLeuAlaProProHisProMetGln 244
Qy 290 CCCGAGGACCCAGAGAGGAGGAGAGTGTGAACCATGTCTTAAGGACATGCTTAAGGACATGAGCTG 349
Db 245 ProValSerValAsnArgGlnMetAsnProAlaAsn----- 256
Qy 350 TTCATGGAAAGCTGGAGAGGCCAGC-----AAAGACCAGCAGGAA 393
Db 257 --PheProGlnLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 275
Qy 394 GAAGAAATTTGGGAAAAAACAAGGACCGAGGTCTCACCCAGGACAGTACATTGA 453
Db 276 GlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 290
Qy 454 CTGCTTCAGAGATCAAGTACAGCTTCAACCTCTCTGGAAGGCTGCGCCACCTGCTGAA 513
Db 291 -----ProGlnGlnHisGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 306
Qy 514 GGAGACAAGTCCCTGAGCTCGTACACATCTCTTCAAGTCCCTGAACTTTCATCCTGCG 573
Db 307 -----ThrAlaProThrGlnValPro----- 313
Qy 574 CAGGTGCCCTGAGCTGCGCTAGCAGCCCAAGTGTAT---CTCACCCCTCTTCACCCCTAA 630
Db 314 --ValProProGlyTrpAsnGlnLeuProSerGlyAlaLeuGlnProPro----- 329
Qy 631 AGCTATCAACCTGTACAGTCTCTGTAAAGCCACCTCAGAGTACCTTTGGATGGGTT 690
Db 330 -----ProAla-GlnGlySerLeuGlyThrMetThrAlaAsnGln-GlyTrp----- 344
Qy 691 GGGCCCTGGACCTAGCCGGGGCCGACTGGACAGCGGATGAGCCCTGCCCT----- 746
Db 345 ----LysLysAlaProLeuProGlyProMetGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 363
Qy 746 ----- 746
Db 363 euAlaThrValGlnThrProSerHisProProProTyProPheGlySerGlnGln 383
Qy 747 -----ACCAAC---CCACATTTCTCAGATCAGTGGCAACTTCAGAGCCCTC 789
Db 383 laSerGlnAlaHisThrAsnPheProGlnMetSerAsnProGlyGlnPheThrAlaPro- 402
Qy 790 CAGCCAAGCACCCTT-----AGGATACACAGGACCTGTTCCTTCCTGGCG 834
Db 403 GlnMetLysSerLeuGlnGlyGlyProSerArgValProThrProLeu----- 418
Qy 835 GGAAGTCATAGTTAGGAGCACCCTCACA-----CTTCTCAGGAGAGAC 882
Db 419 -----GlnGlnProHisLeuThrAsnLysSerProAlaSerProSerSer 434
Qy 883 ACACAACCATGACCTCAGCTGGGGACCCCAACTCCAGGCGCTCCAGCCCAACCTTCG 942
Db 435 PheGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 942
Qy 943 CCAGCCAGCCCTGAAATGCAAGTCTTGTACAGTTTGAAGTAGGACCCAGGGAAGT 1002

